THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume XXX May, 1939 Number 5

STUDIES IN THE PSYCHOLOGY OF MEMORIZING PIANO MUSIC. I. A COMPARISON OF THE UNILATERAL AND THE COÖRDINATED APPROACHES

GRACE RUBIN-RABSON

New York City

INTRODUCTION

In most kinds of learning, the differences between the gifted and the non-gifted seem to be quantitative rather than qualitative. This may be assumed to be true for the acquisition of musical skills, though no experimental work exists to authenticate the assumption. Creative talent might be considered indubitable proof of musical capacity, but, short of this, no reliable tests have been devised which can evaluate the constituent aspects of greater or less "musicality." Virtuosity either on an instrument or with the voice may be notoriously misleading as a gauge of basic abilities. Assiduity, a fine motor equipment, and good training may account for success, as a lack of any one or all of these may account for failure.

Regardless of talent, however, many music students are handicapped by an ignorance of learning techniques. For the most part, verifiable and authoritative learning techniques do not exist. At best, teachers and students alike have been dependent on empirically useful procedures when they have been fortunate enough to discover them. Broadly speaking, good learning techniques must include good memorizing and, hence, good retention techniques. What these techniques are is the scope of this study.[17]

Because of its widespread study and because it includes every practical aspect of the problem, the memorizing and retention of piano music offers the most fruitful area for investigation.

THE PROBLEM

Comparatively few studies have been done in the psychology of music, and of these only a handful deal with the learning or memorizing of piano music. Eberly[4] found advantages in whole over part memorizing, but separated the learning and memorizing functions. This was similar to findings by Brown[2] where the material was not memorized but brought only to smooth performance with the music. Kovacs[12] found great superiority in the retention of musical fragments which were completely learned away from the keyboard as compared to fragments learned directly at the keyboard. In a study[17] testing a similar hypothesis, the writer secured comparable results. The hypothesis was modified, however, by the following assumptions: (1) Only the most gifted can learn music completely away from the keyboard; (2) the value of the pre-keyboard study itself may be negligible because of inadequate study skills; (3) for most students only small fragments can be brought to a point of such learning that the first keyboard repetition is successful; (4) longer experimental materials are indicated to approximate a real learning situation. Therefore, short but entire compositions were studied away from the keyboard for a fixed number of minutes before completing the learning at the keyboard.

Because of the possible inadequacy of study-skills, one of the experimental procedures included the study of the composition with the aid of a prepared outline, contrasted with procedures in which the subject prepared his own outline or used no study period at all and confined the learning exclusively to the keyboard. The results were outstandingly in favor of the pre-study method over the method in which no study period was included; even the inexpert outlines made by the subjects proved much superior to learning directly at the keyboard. Brown,[3] in her investigations in the essentials of learning piano music, finds that practising both hands together is more efficient and more enjoyable than practising either hand alone. Like her other study, the criterion of learning was not memorizing, but simply smooth performance with the music.

The present experiment deals, in a measure, with the same question of unilateral *versus* coördinated learning, but extends the inquiry into the efficiency of either method for memorizing and retention. The question of enjoyability by either method seems to be largely a matter of training and working habits, and, for the serious music student, the

affective factors may be assumed to be a function of the efficiency of the method.

In addition to the main purpose of the inquiry, a subordinate issue is raised. That the study preliminary to actual keyboard practice is superior for retention has already been suggested, since repetitions alone do not in and of themselves produce dependable learning, as was pointed out unequivocally by Thorndike[19] in a series of experiments designed to answer whether or not mere repetition of a situation produces learning.

It must be borne in mind that musical material, unlike verbal, must produce two kinds of meaningfulness, one kind due to an appreciation of the structure and the interrelationships of all the parts, and the other to an awareness of the actual tonal effect. While the first is readily achieved by most students with even a minimum theoretical training, the second is absent in an incredibly large number of even experienced pianists. It might be assumed that the awareness of the tonal effect would greatly aid the analytical study. In a previous experiment,[17] the addition of this element showed no appreciable differences in either learning or retention. This was surely not due to the fact that the subjects were all able to achieve the tonal effect without outside aid, since many of them confessed an inability to do so.

The experiment herein reported, in an attempt again to test the efficacy of the awareness of tonal effect in the analytical study, alters the emphasis somewhat. While previously the subject passively followed the musical score during several recorded repetitions, in the present instance he himself produced several repetitions at the keyboard. This naturally adds to the tonal picture the extra factor of partially formulating the hand patterns.

The major portion of the inquiry is designed to test the null hypothesis that no differences exist in retention efficiency between materials learned unilaterally preceding the coördinated trials and materials learned by coördinated trials alone. The patterns of a right-handed culture combined with the greater audibility of higher as against lower tones have operated to stress the emphasis and training of the right hand in piano study. With rare exceptions, the left hand remains an auxiliary both in performance and in consciousness. Unfortunately, in public performance emotional disturbances alter somewhat the balances of attention, and the inadequacy of the preparation of the left hand as an entity is painfully apparent.

Innately, the structure of piano music is both vertical at the moment when the right and left hands come together, and horizontal as either hand progresses alone. The melody and the harmony grow together and are one. It is possible, then, to apprehend not only the interrelationships of the right and left hands, but either one as an independent configuration. Though in most instances the left-hand musical patterns have not been designed to conform to the criteria of good melodic structure, nevertheless almost any sequence of tones can be perceived as an entity. This is supported by Von Ehrenfels[23]: "The decisive step in the founding of a theory of Gestalt quality was my own assertion: When the memory images of successive notes are present as a simultaneous complex in consciousness, then an idea (Vorstellung) belonging to a new category, can arise in consciousness, a unitary idea, which is connected in a manner peculiar to itself with the ideas of the complex of notes involved."

Broadly speaking, then, three configurations exist, those in either hand alone, and that which results from the integration of the two. Wheeler[24] expresses this Gestalt principle explicitly in the observation that "any item of reality is in its own right an integrated whole that is more than the sum of its parts; it possesses properties not characteristic of its parts. Strictly speaking, the parts are not integrated; it is the whole as such that exhibits integration." Further,[25] "Parts that are relatively segregable and therefore relatively independent are, by definition, units in their own right and follow the laws of wholes; but parts are never absolutely independent no matter how specialized." Restated in these terms the problem is as follows: To what extent does the establishment of independent right- and left-hand configurations aid or inhibit the learning of the integrated configuration; to what extent does it affect retention; to what extent does it establish measurably better learning? Further, do several interpolated keyboard repetitions aid or inhibit the intensity of the pre-keyboard analysis as a preparation for the keyboard learning; does it affect retention; does it insure measurably better learning?

Implicit in these aspects of the problem are four variables which are combined in four learning procedures:

(*A*) The material is studied intensively without interruption before completing the learning by coördinated trials at the keyboard to the level of established criteria.

(*B*) The material is studied intensively without interruption before completing the learning at the keyboard, first by unilateral trials alter-

nated to the point of one smooth memorized performance, succeeded by coördinated trials to the level of established criteria.

(*C*) The study period is interrupted midway for three coördinated keyboard trials. At the end of the study period, the learning is completed at the keyboard by coördinated trials.

(*D*) The study period is interrupted midway for three keyboard trials, one right hand, one left hand, and one coördinated. At the end of the study period, the learning is completed at the keyboard, first bringing the unilateral learning to smooth memorized performance succeeded by coördinated trials.

The four methods provide combinations of Interrupted and Uninterrupted study periods, Coördinated and Separate hand practising. For easy reference, *A*, *B*, *C*, and *D* will be called methods *UC*, *US*, *IC*, *IS*, indicating, respectively, uninterrupted and coördinated, uninterrupted and separate, interrupted and coördinated, interrupted and separate hand approaches.

EXPERIMENTAL MATERIALS

Four complete eight measure fragments were chosen for their simplicity, attractiveness, and unfamiliarity. Only such music was considered for the purpose as contained left-hand structures of some individuality and independence. Simple chord basses occurring only once or twice in a measure or repeating several times without change and which are inherently only a repetition or completion of the melodic line above were not considered suitable.

The fragments used were adapted from:

1. Frescobaldi (1583–1643) Canzone
2. Haydn (1732–1809) Piano Trio
3. Haydn Piano Trio
4. Butting (1888–) Allegretto

The four compositions represent a considerable range in era, in style, and in melodic appeal. Though it is impossible to find four compositions which all subjects would consider of equal difficulty, these are all simple enough to be handled readily by subjects with not less than five years of piano training. Furthermore, though simple, they offer a fair cross section of much of the material in the piano repertory.

Thirty-six subjects were chosen from the music schools of New York City. All of them were serious music students or professional musicians. None had had less than five years of formal piano instruction or

less than two semesters or the equivalent of theoretical training. The group ranged from this minimum to seventeen years of piano training and twenty-four semesters of theoretical training. Ages ranged from eighteen to twenty-six years. There were thirty women and six men. The piano and theoretical training figures were derived during individual interviews with the subjects and evaluated as uniformly as possible from the professional experience of the experimenter.

Table I indicates some of the attributes of the experimental group as a whole.

TABLE I.—MEANS, STANDARD DEVIATIONS, AND RANGE OF AGE AND MUSICAL EXPERIENCE; MEANS AND RANGE OF ACHIEVEMENT FOR THE EXPERIMENTAL GROUP AS A WHOLE

	Means	Standard deviation of the distribution	Range
Age	22.0	2.4	18 –26
Piano experience (years)	9.75	2.8	5 –17
Theoretical training (terms)	10.0	5.8	2 –24
Learning trials	14.0	...	3.3–27.3
Relearning trials	10.8	...	3.5–24.8
Transcription score errors	19.1	...	1.8–53.8

The comparatively small standard deviation shows the subjects to be fairly homogeneous as to age and at an average of twenty-two years, at approximately the optimum learning age, according to Thorndike.[20] The average of nearly ten years of piano study with small divergence indicates the persistence of the subjects as serious music students. An average of ten terms of theoretical training points to the same conclusion, but the variations from the mean are greater.

EXPERIMENTAL DESIGN

In addition to the four learning procedures, it is necessary to use an experimental design in which these variables are rotated, or presented to each group of subjects in all possible orders and combinations so that their influence is distributed equally in the four learning methods. The Latin-Greek square offers the most precise experimental design for the analysis of the efficiency of four treatments or procedures when individual differences are known to exist but can not be controlled. When the thirty-six subjects are divided into four groups, the precision is increased by the nine-fold replication of each treatment. The

known variables affect each treatment equally and the proportion of unknown factors can be segregated by the analysis of the variance. For each group of subjects, the variables of treatment (procedure), experimental material and order are combined differently and in rotation. The analysis, therefore, will include these four variables: Group differences, procedures or methods, compositions, and order of learning.

The efficiency of the treatments can be evaluated by the number of trials required to bring performance to an established level of proficiency; by the difference between trials needed to restore the proficiency after a lapse of time; by a comparison of trials required by each of the four procedures for the relearning; by some arbitrary analysis of the precision and clarity of the learning itself.

METHODOLOGY*

All learnings were done individually, with only the subject and the experimenter present. Two learnings were accomplished at each session with a short rest period between them. The relearnings were carried through in the same order exactly two weeks from the date of the learning, and followed the same experimental pattern as the learnings except that the preliminary study period was omitted and all the work was confined to the keyboard and the relearning done by the coördinated approach.

Each step in the experimental procedure was clocked to the second. Time for comfortably reading directions and going to and from the piano were not included in the prescribed time limits. Each trial of the material was checked at its termination.

The length of the study period was the same to the second for each learning. Where the study period was interpolated by keyboard trials, the mid-point was at exactly seven minutes. The length of the three keyboard trials was held at constant metronomic speed. The duration of the three trials for each fragment is as follows:

1. Seventy-five seconds, M. M. ♩—116
2. Forty-eight seconds, M. M. ♩—92
3. Sixty-three seconds, M. M. ♩—92
4. Forty-eight seconds, M. M. ♪—92

When the study period was uninterrupted, the number of seconds for the fragment under consideration was added to the original fourteen minutes.

* The experiment was conducted in the studio of the experimenter in New York City between June 21 and September 10, 1937.

At the end of each session the subject was admonished against practising the material at home, and was urged, if possible, not to think of it again.

After achieving the criterion of relearning two weeks later, two techniques were employed to gauge the quality of the learning. The subject was asked to play each hand separately, a detailed record being kept of his success. With this preliminary preparation, he then transcribed the score from memory as accurately as possible on music manuscript paper.

Implicit in the methodology are certain assumptions and experimentally authenticated facts:

(1) With direction and sufficiently intensive attention, the subject can achieve some awareness of the tonal effect of the material without playing it. This capacity is undeveloped in a surprisingly large number of piano students who, though they may have played common intervals and tonal combinations literally thousands of times, can secure no tonal images on seeing the same symbols away from the keyboard. Though this capacity would seem to be an inevitable concomitant of piano study, in most cases it reaches development only through special ear-training and sight-singing techniques.

(2) Attention is directed not only to the relationships of the right and left hand as an integrated pattern, but to the right and left hands as separate configurations. Few piano students concern themselves with analysis of the left hand as an entity and give it special attention only when specific technical or coördination problems arise.

(3) The groupings and organization of the material are not only observed but verbalized.

(4) The learning and memorizing are treated as one function. True learning may be said to exist only when the material has been meaningfully memorized, since a flawless rendition may be executed by a skillful pianist at first sight. The will to memorize demands at once an intensity of attention almost never achieved when the memorizing is superimposed as a separate and discrete function. The importance of this intensity is emphasized by Hughes[10]: "Retentiveness depends largely on the strength of the impressions made during the first attempts to master the composition."

(5) The performance of memorized piano material seems to combine real memory with the operation of kinaesthetic habit built up of constant repetition of finger patterns. If methods of learning are to be adequately evaluated for their proficiency in producing real memory, it

is necessary to test the finished performance for the proportion of sheer habit entering into it.

(6) Stress is laid on grouping the musical materials into patterns, repetitions, transpositions, chord lines, and the like. Grouping and organization of the material whether it be meaningful or non-meaningful, are the decisive factors in the learning and memorization. Martin and Fernberger[15] found grouping and organization imperative in learning longer series and the capacity for grouping improvable. Vogelsang[22] finds that the ability to group depends not only on the subject but on the character of the material; also that better subjects are able to deal with larger groups than poorer ones. This is corroborated by Ewald.[5]

Musical materials must be considered as intrinsically meaningful. The groupings are, indeed, facilitated by constant return to a tonal center as a point of rest as well as by the quite natural organization into rhythmic groups. Elkin[6] points to the readier availability of rhythmic material.

(7) The learning is accomplished by the whole method insofar as an eight measure fragment may be considered a complete musical unit, and was repeated each time without interruption. The superior results which Eberly[4] found for the whole method must be accepted with reservations. The whole method was adopted in the present instance in order to control the data more accurately.

(8) Errors in note, rhythm, and coördination were not allowed, but were checked and corrected immediately on occurrence. The persistence of first wrong impressions has been stressed by Thorndike.[19]

(9) Prompt performance without the score utilizes the principle enunciated by Gates[8] that "only a very small percentage of the total time required to learn should be devoted to reading."

(10) Because of experimental exigencies the learning was done by massed practice. As a learning technique it has been found inferior to distributed practice by Jersild,[11] Lorge,[13] and Trow.[21] McGeogh[16] sums up the results of numerous studies in the law that "some form of distribution is always superior to massed practises. It holds for virtually all experimental techniques and for all subjects and materials."

ANALYSIS OF DATA

Two basic statistical techniques have been employed to evaluate the relative efficiency of the four experimental procedures. An analysis of the variance according to Fisher's[7] formula, in conjunction with

the test for the significance of the differences of the means, establishes reliably not only the differences between the methods themselves but also how significant a factor they are when compared with the other variables entering into the experimental situation.

Table II shows the variation among the experimental compositions as indicated by the average number of trials needed for learning and relearning and the errors in the transcription score.

TABLE II.—DIFFERENCES IN THE FOUR EXPERIMENTAL COMPOSITIONS AS INDICATED BY THE AVERAGE LEARNING AND RELEARNING TRIALS AND BY THE AVERAGE ERRORS IN THE TRANSCRIPTION SCORE

	Learning trials	Relearning trials	Relearning per cent of learning	Transcription score
Composition 1.....	21.7	14.4	66	29.2
Composition 2.....	16.8	10.9	64	21.5
Composition 3.....	11.1	8.5	77	13.9
Composition 4.....	16.0	9.4	59	11.7

LEARNING TRIALS

Table III shows the analysis of the variance of the learning trials. For the procedures in which unilateral preceded coördinated practice, the number of unilateral trials was divided by two and added to the coördinated trials. It is not to be inferred that two unilateral trials have the precise learning value of one coördinated trial. As a matter of fact, this was found not to be true. For the purposes of comparison, however, it may be assumed that one coördinated reading includes a reading of either hand, as two unilateral trials includes one reading of either hand.

Only compositions, accounting for eighty-six per cent of the total variance, exceeds Fisher's[7] value, a value which might be exceeded in random sampling from a homogeneous population five times in a hundred trials. It is also the only variable which begins to approach 29.46, Fisher's one per cent value which might be exceeded in random sampling once in a hundred trials. Reference to Table II, showing the extreme differences among the compositions, confirms this point. The experimental group proved fairly homogeneous, accounting for only five and four-tenths per cent of the variance. That the subjects quickly adapted themselves to an altogether novel learning technique, and that the learning procedures exerted very little influence on each

other is borne out by the three per cent contribution of order. Uncontrolled factors or the residual error accounted for only four and six-tenths per cent, a satisfactorily small proportion. The methods themselves contributed only one per cent, a negligible fraction.

TABLE III.—ANALYSIS OF THE VARIANCE OF THE LEARNING TRIALS, RELEARNING TRIALS, AND TRANSCRIPTION ERRORS SHOWING THE PERCENTAGE OF THE TOTAL DUE TO EACH VARIABLE, AND A COMPARISON OF THE RATIO OF THE VARIABLE AND THE EXPERIMENTAL ERROR WITH FISHER'S FIVE PER CENT VALUE

	Ratio of variable and experimental error	Fisher's five per cent value	Per cent of whole
Learning			
Groups	1.36	9.28	5.4
Methods	.26	9.28	1.0
Compositions	21.50	9.28	86.0
Order	.75	9.28	3.0
Error	.....	9.28	4.6
Relearning			
Groups	.25	9.28	1.6
Methods	.38	9.28	2.4
Compositions	13.30	9.28	84.0
Order	.82	9.28	5.2
Error	.....	....	6.8
Transcription Score			
Groups	2.6	9.28	7.3
Methods	8.3	9.28	23.5
Compositions	17.9	9.28	50.7
Order	5.5	9.28	15.6
Error	.....	9.28	2.9

Table IV shows the means and standard deviations of the methods. For each method the number of coördinated trials appears, as well as the total of the coördinated trials plus the unilateral trials. Comparison of the coördinated trials alone shows a considerable difference among the methods. Comparison of the total trials shows very small differences among the methods. The two coördinated methods,

whether preceded by the interrupted or by the uninterrupted study period, show remarkable conformity. The two unilateral methods show a difference of only one and two-tenths trials in favor of the procedure which incorporated keyboard trials during the study period. For the coördinated learning, the three preliminary trials during the study period reduce the final learning average by only four-tenths of a trial. For the unilateral procedures the right hand required six and eight-tenths trials, regardless of the preliminary trials, in the *IS* method. The left hand, less stressed in training, required a greater number of trials, whether *US* or *IS*, the difference of one trial between them representing, perhaps, the one left-hand preliminary trial in the *IS* procedure. For the most part, the deviations from the mean seem to be proportionally greater in both interrupted methods, whether unilateral or coördinated.

TABLE IV.—MEANS AND STANDARD DEVIATIONS DERIVED FROM THE LEARNING TRIALS FOR THE FOUR EXPERIMENTAL METHODS

	Mean	SD of distribution
UC	16.5	8.6
US		
Right hand alone	6.8	3.4
Left hand alone	8.1	5.4
Coördinated Trials	12.1	6.2
Total: Coördinated and Unilateral Trials	19.5	10 8
IC	16.1	9.8
IS		
Right hand alone	6.8	4.8
Left hand alone	7.1	3.8
Coördinated Trials	11.4	7.0
Total: Coördinated and Unilateral Trials	18.3	11.6

Table V presents the same data with somewhat different emphasis. Here the like methods have been averaged. The two uninterrupted methods, the two interrupted methods, the two unilateral methods, and the two coördinated methods are compared for the number of coördinated trials required to reach the learning criteria, as well as the coördinated trials plus the unilateral trials. Further, the preliminary trials have been added to form a total in the interrupted methods. The uninterrupted *versus* interrupted procedures show negligible differences, the largest difference of one and seven-tenths trials occurring

only when to the interpolated procedures a total of five preliminary trials has been added. It may, therefore, be inferred that no advantage accrues to the interpolated study period. The unilateral *versus* coördinated methods naturally show greater differences. Comparing only coördinated trials, a difference of three and six-tenths trials exists in favor of the separate hand approach (Table V). Comparing total learning trials, the advantage accrues to the coördinated approach by two and six-tenths trials. When all trials, including preliminary trials, are compared, the difference is smallest, being two and one-tenth trials in favor of the coördinated procedures. Table VI shows the significance of these differences. Considering only coördinated trials, whether interrupted or uninterrupted, the coördinated methods show practically no differences. This is also true for the separate hand methods. Comparison of the coördinated *versus* the separate hand methods, however, indicates considerable differences, the ratio of the differences to the standard error being in every case two or more. The most reliable relation, 2.69, exists between the *UC* and the *IS* methods, favoring the latter. That this difference is not due to the *U* and *I* aspects is indicated by the negligible differences when only these aspects are compared.

Comparing the total numbers of trials required by the four procedures, some differences exist, but they are not statistically reliable. Again, when only the *U* and *I* aspects are considered, the differences are negligible. The separate and coördinated methods show greater differences, but can not be considered significant.

Reference to Table V shows a difference of three and six-tenths trials between the *S* and the *C* procedures. Table IV presents the data of the right and left hand trials required before the coördinated trials in the *S* methods. The average of these left- and right-hand trials together is fourteen and four-tenths. Stated differently, four separate hand trials were required to effect a saving of one coördinated trial, instead of the two separate hand trials which might have been expected.

Two hypotheses may be advanced in explanation of the comparatively large number of separate hand trials required to reduce the total by one coördinated trial. One is that transfer in the learning undoubtedly exists, but it is both positive and negative. It is possible that the previous training of the subjects accounts for the large negative transfer and that specific training might tend to eliminate it. An alteration in the experimental technique might also have effected a radical change

in these results. It will be remembered that the unilateral learning was brought to memorization before the learning was completed by coördinated trials. Had one coördinated trial been interpolated between each two unilateral trials, much of the negative interference might have been eliminated and more real differences become apparent

TABLE V.—COMPARISON OF LEARNING TRIALS FOR THE FOUR EXPERIMENTAL METHODS WHEN LIKE METHODS ARE GROUPED

	UC, US	*IC, IS*	Difference	*US, IS*	*UC, IC*	Difference
Average coördinated trials	14.3	13.7	0.6	12.7	16.3	3.6
Average coördinated trials plus unilateral trials....	18.0	17.2	0.8	18.9	16.3	2.6
Average coördinated trials, unilateral, and preliminary trials.....	18.0	19.7	1.7	19.9	17.8	2.1

between the methods. Insofar as the learning is a habit skill built up in either hand by several repetitions, these findings receive support by Beeby[1] that, in tracing skills, both positive transfer and negative interference appear in the transfer from single- to double-handed practice.

The second hypothesis is that the configuration built up in either hand is only a peripheral configuration, while the coördinated configuration is the dominant one. This might be restated in Wheeler's[26] law of learning that "wholes evolve as wholes in the learning and that at every step the whole is there."

On the whole, no economy is evidenced in the learning trials by any of the four methods. Statistically, it seems immaterial whether the study period preliminary to the keyboard learning is interpolated by several keyboard trials in order to secure a tonal image as an aid to further non-keyboard study, or whether the pre-keyboard study period is uninterrupted, although there is some slight evidence in favor of the latter. Though some negative interference seems to operate in the transition from the separate hand to coördinated trials, some positive transfer exists as well, and the difference between the two procedures indicates superiority in neither one.

RELEARNING TRIALS

The analysis of the variance of the relearning trials (Table III), though indicating some slight shifts in proportional percentages,

TABLE VI.—SIGNIFICANCE OF THE DIFFERENCES OF THE MEANS DERIVED FROM THE LEARNING TRIALS OF THE FOUR EXPERIMENTAL METHODS (1) WHEN ONLY COÖRDINATED TRIALS ARE COMPARED; (2) WHEN THE TOTAL NUMBER OF TRIALS IS COMPARED

(1)

Methods	Difference	SD difference	Difference to SD difference
UC-US	4.4	1.8	−2.44
UC-IC	.4	2.2	−0.18
UC-IS	5.1	1.9	−2.69
US-IC	4.0	1.9	2.10
US-IS	.7	1.6	−0.43
IC-IS	4.7	2.3	−2.04

(2)

UC-US	3.0	2.3	1.30
UC-IC	.4	2.2	−0.18
UC-IS	1.8	2.4	0.75
US-IC	3.4	2.5	1.36
US-IS	1.2	2.7	−0.44
IC-IS	2.2	2.6	0.85

presents practically the same situation as the variance of the learning trials. Compositions remain the largest and most significant factor, accounting for eighty-four per cent of the total. Methods, though somewhat more important than before, remain small, contributing two and four-tenths per cent. The differences between the groups is reduced, being only one and six-tenths of the whole. Order and the residual account for slightly higher percentages, being five and two-tenths per cent and six and eight-tenths per cent, respectively. Of the five component factors, only compositions exceeds Fisher's five per cent value.

The means of the methods given in Table VII bear out the comparative insignificance of methods in the variance of the relearning trials. The number of trials required to relearn four compositions learned two weeks earlier, each by a different approach, achieve a remarkable uniformity, deviating among themselves by no more than one trial. Small as are the differences, any possible advantage accrues to the uninterrupted study period procedures, both the *UC* and the *US* methods learning to the smaller means. These differences are cited only as possible tendencies, since Table VIII shows them all to be statistically insignificant.

The conclusion seems inescapable that, in this instance, no learning method offers any advantage in the amount of retention as measured by the relearning trials. This finding offers repeated confirmation of Wheeler's[24] observation that "the parts are not integrated; it is the whole as such that exhibits integration." Whatever the early procedures in the original learning, they seem to recede in consciousness as the successive procedures establish dominance.

TABLE VII.—MEANS AND STANDARD DEVIATIONS DERIVED FROM THE RELEARNING TRIALS OF THE FOUR EXPERIMENTAL METHODS

	Mean	SD of distribution
UC	10.8	7.2
US	10.3	6.0
IC	11.3	6.2
IS	10.9	5.2

TABLE VIII.—SIGNIFICANCE OF THE DIFFERENCES OF THE MEANS OF THE RELEARNING TRIALS OF THE FOUR EXPERIMENTAL METHODS

Methods	Difference	SD of difference	Difference to SD difference
UC-US	.5	1.6	−0.31
UC-IC	.5	1.6	0.31
UC-IS	.1	1.5	0.06
US-IC	1.0	1.4	0.71
US-IS	.6	1.3	0.46
IC-IS	.4	1.4	−0.28

PRECISION AND CLARITY

Neither the learning trials nor the relearning trials indicate any superiority in the methods. As judged by number of trials and fluency of performance, the efficiency of the learning is the same. Beneath the fluency, however, perhaps some measurable differences exist in the precision and clarity of the learning.

To test this precision two tests were made. When the relearning was completed, the subject was required to play each hand alone from memory. The results, by methods, appear in Table IX. The greatest success for the right hand alone derives from the *US* method, in which twenty-eight subjects produced perfect performances. Next, with twenty-two, is the *IS* method; third with nineteen is the *IC* method,

and fourth with seventeen is the *UC* method. Both unilateral methods produce superior right-hand performances. Though the differences are smaller, the same superiorities exist for the left hand by the unilateral methods.

Although not perfectly smooth, many subjects obtained performances which were without errors or omissions but which were, nevertheless, hesitant. These were divided into two classes—those with "some hesitation" and those with "much hesitation." In the "some hesitation" group, the interpolated methods totaled fifteen right-hand performances and twenty-four left-hand performances, while the coördinated methods totaled twelve right-hand performances and six left-hand performances. The "much hesitation" group shows one right-hand and nine left-hand performances for the unilateral methods and three right-hand and five left-hand performances for the coördinated methods. Altogether, the unilateral methods produced sixty-six right-hand and fifty-seven left-hand performances without errors or omissions, the coördinated methods, fifty-one right-hand and twenty-four left-hand.

Measures omitted or distorted totaled ten and one-half right-hand and forty and one-half left-hand for the unilateral methods and forty-nine and one-half right-hand and one hundred seventy-three and one-half left-hand for the coördinated methods. For security of learning as demonstrated by the ability to perform each hand alone, the unilateral methods show definite superiority. Because of the extreme simplicity of the experimental material, the preliminary study period undoubtedly guaranteed perfect single-hand performances for the abler subjects even when the actual keyboard trials were all of the coördinated type.

If, as Hollingworth says,[9] the auditory cues are adequate to accomplish the result, (*i.e.* memorizing piano music) what is left of the learning when the auditory cues are removed? If, as Mainwaring,[4] says the memorizing is largely kinaesthetic, what is left of the learning when the overt kinaesthesis is suppressed? To determine this quantitatively, the subject, after each relearning, transcribed the experimental material on music paper. The scoring of omissions, additions, errors, transpositions, and the like was standardized for all the compositions, hence the lower the score the better the transcription.

Table III presents the analysis of the variance of the transcription score. Though compositions is still the only variable producing a ratio which exceeds Fisher's five per cent value, it has decreased from the

former eighty-six per cent and eighty-four per cent for the learning and relearning variance to fifty and seven-tenths per cent in the present instance. This is the more striking as the more difficult compositions by their very structure incurred disproportionately heavier penalties in the scoring. The increase of order to fifteen and six-tenths per cent of the whole might be due to the emotional adaptation of the subjects, many of whom felt somewhat nonplussed on first being asked to write the transcription. The group variation shows only slight change from the variation in the learning variance.

TABLE IX.—PERFORMANCE OF LEFT AND RIGHT HANDS ALONE AFTER RELEARNING, SHOWING NUMBER OF PERFECT AND HESITANT PERFORMANCES AND NUMBER OF MEASURES OMITTED

	UC		*US*		*IC*		*IS*	
	Right hand	Left hand	Right hand	Left hand	Right hand	Left hand	Right hand	Left hand
Perfect	17	10	28	13	19	3	22	11
Some hesitation	7	1	6	13	5	5	9	11
Much hesitation	1	2	0	3	2	3	1	6
Total	25	13	34	29	26	11	32	28
Measures omitted								
½	2	2	2	2	1	2	2	
1	1	2	..	2	....	1	....	1
1½	2	..	2	..	....	2		
2	1	..	..	1	2			
2½	1	5	..	..	....	2	....	2
3	2	3	..	1	1	1	1	1
3½	1	1	..	1	2	1	1	
4	..	3	..	..	1	4	....	2
4½	..	2	..	..	1	2	....	1
5	..	..	..	..	....	2	....	1
5½	..	1	..	..	....	1		
6	1	1	..	..	....	5		
6½	..	..	..	..	....	1		
7								
7½	...	1						
Total	25	80	3	14	24½	93½	7½	26½

Methods, however, takes on a new significance, contributing twenty-three and five-tenths per cent of the whole. For the first time, methods closely approaches Fisher's significant five per cent value. This proportion is confirmed by Table X where the means of the

transcription errors produced by the various methods is seen to conform to a distinct pattern, the means of the coördinated methods lying close together, as do the means of the unilateral methods, the deviations of all four being extremely large in proportion to the means.

Table XI shows the largest and most reliable differences between *US* and *IC* methods and between the *UC* and *US* methods. In both differences the advantage lies with the unilateral method. It is interesting to observe that the more reliable difference lies between the methods of which one used the interrupted study period. The means of the coördinated methods, as also the means of the unilateral methods, show small differences. Real differences seem to exist between the means of the *UC* and *IS* and between the *IC* and *IS* methods, but the ratios of the difference of the mean and the standard error lie below the conventional significant ratio.

Though no one method evidences any superiority in learning or relearning trials, nevertheless, both unilateral methods offer distinct superiority in the clarity of retention and the vividness of the images after both tonal and kinaesthetic stimuli have been removed. Consistently the methods incorporating the interpolated repetitions at the mid-point of the preliminary study period either add nothing to the learning or exert positive interference with the efficiency of the learning.

CORRELATIONS

Since the age range was limited and a minimum amount of piano and theoretical training was required, and since, in a negative sense, the musical training of the group was similar in that the emphasis on efficient learning procedures had been neglected, a certain homogeneity tended to keep the correlations of various factors somewhat smaller than they might otherwise have been.

The correlations presented in Table XII provide some interesting contradictions to the belief common among music educators that long years of memorizing piano music aided by intensive theoretical training improve the memorizing capacity. The number of learning and relearning trials show a dependence of only -0.284 and -0.220 on piano experience.

Theoretical training shows correlations of only -0.244 and -0.265 with the learning and relearning trials. This may be due in part to a fairly high degree of correlation, 0.470, between the common factors of piano and theoretical training. More likely, however, are the following: (1) Few piano students utilize their theoretical training in the

analysis of piano music; (2) a small amount of theoretical training is adequate to group the material efficiently and that, past this point, the greater experience does not show proportional efficiencies; (3) more transfer can be expected from training in specific memorizing skills than from a general background of form, structure, and chord progressions.

In view of the fairly high correlation between piano and theoretical experience, it seems expedient to partial them out of the other relationships. When theoretical training is held independent for the group, the relationship between piano experience and learning trials is increased to −0.371, although the difference does not warrant much change in the previous conclusion. When piano experience is assumed to be the same, the degree of association between theoretical training and learning trials drops to −0.132, lending increased weight to the foregoing suppositions.

A correlation of −0.227 between the amount of piano experience and the transcription score is low, being reduced even further to −0.157 when theoretical training is isolated. It seems logical to infer that if long piano experience in itself provides no techniques for forming dependable mental or finger images, it will not form images clear enough to set down accurately on paper. The correlation between

TABLE X.—MEANS AND STANDARD DEVIATIONS OF THE ERRORS OF THE TRANSCRIPTION SCORES BY THE FOUR EXPERIMENTAL METHODS

	Mean	SD of distribution
UC	23.6	22.6
US	13.3	13.6
IC	23.7	19.2
IS	15.7	16.4

TABLE XI.—SIGNIFICANCE OF THE DIFFERENCES OF THE MEANS DERIVED FROM THE TRANSCRIPTION SCORES OF THE FOUR EXPERIMENTAL METHODS

Methods	Difference	SD of difference	SD of difference to difference
UC-US	10.3	4.3	2.39
UC-IC	0.1	4.8	0.02
UC-IS	7.9	4.6	1.72
US-IC	10.4	3.9	2.67
US-IS	2.4	3.6	0.67
IC-IS	8.0	4.2	1.91

theoretical training and the transcription score is again low, −0.282, and is reduced to −0.204 when piano training is held constant. This latter low correlation is the more unusual, since theoretical training stresses not only the writing of chords and correct rhythms, but should aid the student in transcribing by solfeggio when his visual imagery fails him. A strange dichotomy becomes apparent in the theoretical training and its practical application.

TABLE XII.—CORRELATIONS OF MUSICAL EXPERIENCE AND EXPERIMENTAL ACHIEVEMENT

	Theoretical training	Learning trials	Relearning trials	Transcription score
Piano experience	0.470	−0.284	−0.220	−0.227
Theoretical training	.....	−0.244	−0.265	−0.282
Learning trials	.....	.....	0.784	0.716
Relearning trials	.....	.....	.....	0.520
Partial correlations				
Piano experience (theoretical training held independent)	.....	−0.371	.....	−0.157
Theoretical training (piano experience held independent)	.....	−0.132	.....	−0.204

The relationship between the ability to organize the material into groups, to hold these groups firmly, to visualize them tonally, and to set them down accurately is demonstrated clearly in the high correlation between learning trials and the transcription score, 0.715, and the relearning trials and transcription score, 0.520. When the learning is prompt, the written score is firm and sure. The efficient learner obviously does not depend on repetition alone for the learning, but on a skillful organization of the material which begins at once and which carries through into the writing. It might be expected that these relationships would have been even higher had the score been written after the learning trials instead of after the relearning trials. It will be recalled that no study period preceded the relearning, and that the original learning was carried only to the one hundred per cent learning point and allowed for no over-learning. Hence the visual images during the learning must have been very stable to resist the ravages of much retroactive inhibition due to the interpolation of similar material during the two week interval.

This is even more forcefully demonstrated in the correlation of 0.784 between the learning and the relearning trials. Many of the poorer learners seem to have returned almost to the starting point, so little difference do they show in the relearning trials. Evidently, a large number of trials without real grasp of the materials produces a learning built of a habit skill, sufficiently firm to satisfy the learning criteria, but so ephemeral as to be readily dissipated after an interval of two weeks.

SUMMARY AND CONCLUSIONS

In the search for optimum skills on a keyboard instrument, the criteria of efficiency must differ somewhat from those which suffice for learning skills in other areas. Performance, though apparently kinaesthetic, is postulated on a high degree of organizational skill in the learning. Without this organization of materials, the learning descends to the level of purely motor skill, depending on countless repetitions for habit formation.

Speed in learning, while undoubtedly correlating highly with efficiency, is superseded as a criterion by the clarity and precision which remain when both the kinaesthetic and the tonal factors are eliminated from the learned materials. The amount of retention achieved by various learning techniques may obscure the true situation because of the tendency of the organism to carry over only the integrated phases of the learning and to suppress those on which the integrated phase is constructed.

The main purpose of the inquiry was to compare the security of learning achieved by an approach in which either hand was memorized before coördination, with an approach in which the learning was coördinated from the beginning. In addition, an unbroken preliminary study period was contrasted with a study period interpolated midway by several keyboard trials.

The total learning trials required by the four methods show no reliable differences. A comparison of the coördinated trials alone, however, indicates that four separate hand trials were required in the unilateral methods to effect a saving of one coördinated trial. Transfer apparently exists, but it must be assumed to be both positive and negative. It is possible that an altered emphasis in the subjects' previous training and some change in the mechanics of the experimental situation would operate to reduce this negative interference considerably.

The relearning trials after an interval of two weeks produced a startling uniformity in results. Retention afforded by the four methods shows no differences. However, the unilateral methods showed decided advantages in the ability to play either hand alone and to transcribe the experimental material on music paper. The extreme simplicity of the compositions strengthens these implications for more difficult material, since the preliminary study period undoubtedly guaranteed good separate hand reproduction and fairly accurate transcriptions to the more able subjects even when all the keyboard trials were coördinated.

Consistently the interrupted study period indicated either no advantage or positive interference with the learning. The grouping and organization of the musical materials seems to proceed more vividly when removed from the coördination problems at the keyboard, even though the factor of tonal awareness may be added when the studying is interrupted by keyboard trials.

Only slight association exists between the amount of piano study and the number of learning trials; even less between the amount of theoretical training and the learning trials. Haphazard techniques do not seem to improve memorizing capacity even when exercised over many years; nor does an increasing amount of theoretical training assure proportional efficiency. Low correlations between piano training and the transcription score, and the theoretical training and the transcription score bear out the same conclusions. The need for specific learning techniques is indicated quite apart from the techniques of piano and theoretical training.

The correlation of .715 between learning trials and the transcription score shows a high relationship between the speed and the accuracy of the organization of the materal. This is confirmed by the .784 correlation between learning and relearning trials. Where the learning required many repetitions and is hence largely a motor performance, nearly the same number of repetitions is required to restore it to the original state of proficiency.

It may safely be concluded that unilateral trials preliminary to coördinated trials operate to produce greater stability and clarity in the learning of piano material, even though this superiority is not apparent in the number of trials required either for learning or for relearning. Further, even though studying the score away from the keyboard produces little or no tonal image, no advantage derives from

interrupting the study period with keyboard trials designed to secure this tonal image.

BIBLIOGRAPHY

1. Beeby, C. E.: "An experimental investigation into the simultaneous constituents in the act of skill." *Brit. Jour. of Psych.*, 1930, Vol. XX, pp. 336–353.
2. Brown, R. A.: "A comparative study of 'whole' 'part' and 'combination' methods of learning piano music." *Jour. Exp. Psych.*, 1928, Vol. XII, pp. 235–248.
3. Brown, R. A.: "The relation between two methods of learning piano music." *Jour. Exp. Psych.*, 1933, Vol. XVI, pp. 435–441.
4. Eberly, L. E.: *Part versus whole method in memorizing piano music.* Unpublished Master's thesis. Columbia University, 1921.
5. Ewald, F.: "Untersuchungen über die Komplexweite des Gedächtnisses." *Archiv F. D. ges. Psych.*, 1929, Vol. LXVII, pp. 161–240.
6. Elkin, D.: "Ueber den Einfluss des Rhythmus und des Tempos auf den Gedächtnisprozess." *Archiv. f. d. ges. Psych.*, 1928, Vol. LXIV, pp. 81–92.
7. Fisher, R. A.: *Statistical Methods for Research Workers.* Oliver and Boyd, Edinburgh, 1934.
8. Gates, A. I.: "Recitation as a factor in memory." *Archives of Psychology.* Vol. VI, No. 40, p. 104.
9. Hollingworth, H. L.: *Psychology: Its Facts and Principles.* D. Appleton and Co., New York, 1928, p. 208.
10. Hughes, E.: "Musical memory in piano playing and piano study." *Musical Quarterly*, October, 1915.
11. Jersild, A. J.: "Primary, Recency, Frequency, Vividness." *Jour. Exp. Psych.*, 1929, Vol. XII, pp. 58–70.
12. Kovacs, S.: "Untersuchungen über das musikalische Gedächtnis." *Zeit. f. ang. Psych.*, 1916, Vol. XI, pp. 113–135.
13. Lorge, I.: *Influence of regularly interpolated time intervals upon subsequent learning.* Teachers College (Columbia University) Contr. to Education, 1930, No. 348, pp. 57.
14. Mainwaring, J.: "Cognitive processes involved in music ability and in musical education." *Brit. Jour. of Ed. Psych.*, Feb. 1936, pp. 100–110.
15. Martin, P. R. and Fernberger, S. W.: "Improvement in memory span." *Amer. Jour. Psych.*, 1929, Vol. XLI, pp. 91–94.
16. McGeogh, J.: Robinson: *Readings in General Psychology.* "Exp. Studies of Memory." Univ. Chicago Press, 1929.
17. Rubin-Rabson, G.: "The influence of analytical pre-study in memorizing piano music." *Archives of Psych.*, 1937, No. 220.
18. Thorndike, E. L.: *Fundamentals of Learning.* Bureau of Publications, Teachers College, Columbia University, 1932.
19. Thorndike, E. L.: *Human Learning.* Century Co., New York, 1931.
20. Thorndike, E. L., Bregman, E. O., Tilton, J. W., and Woodyard, E.: *Adult Learning.* Macmillan, New York, 1928.
21. Trow, W. C.: "Recall *vs.* repetition in the learning of rote and meaningful material." *Amer. Jour. Psych.*, 1928, Vol. XL, pp. 112–116.

22. Vogelsang, H.: "Experimentelle Untersuchungen über die Einprägung simultan gebotener Objekte." *Archiv. f. d. ges. Psych.*, 1929, Vol. LXVII, pp. 257–306.
23. Von Ehrenfels, C.: "On Gestalt Qualities." *Psych. Review*, November, 1937, pp. 521–524.
24. Wheeler, R. N.: *Laws of Human Nature.* D. Appleton and Co., New York, 1932, p. 71.
25. Wheeler, R. N.: *Ibid.*, p. 75.
26. Wheeler, R. N.: *Ibid.*, p. 175.

RESEARCH ON THE INTERVIEWING PROCESS

PERCIVAL M. SYMONDS

Teachers College, Columbia University

The interviewing process is an extremely elusive and unwieldy phenomenon to study scientifically. The term interviewing itself is so poorly defined and is used to describe so many variations of the process of verbal communication between two persons that one must specify the kind of interviewing he has in mind before it is possible to think of studying it as a process.

Commonly, interviewing is thought of as the process by which one person—the interviewer—directly addresses another person—the interviewee—in order to learn from him. The interviewer is concerned with getting facts or opinions which the interviewee can supply, or the interview is used as a method of studying the personality of the interviewee. It is usual to think of the interviewing process as largely oral questioning on the part of the interviewer and oral answers given to the questions by the interviewee. If the purpose of the interview is to know the interviewee better it has been discovered that the question and answer method, in general, is not the best procedure to follow. Persons in general do not reveal themselves immediately and directly in answer to personal questions by a stranger. Before a person will tell intimate things about himself he must know who the interviewer is, must have gained confidence in him, must become convinced that he can trust him, and that he will suffer no personal harm or inconvenience or lowering of prestige as a result of what he reveals.

The study of interviewing as a process has two phases: One, the measurement or description of the results of interviewing in the several ways which seem important; and, two, studying these outcomes in the light of certain variations in the interviewing process itself.

If the interview is thought of as comparable to other testing and personality evaluating instruments it may be studied under the captions *validity* and *reliability*. Indeed, previous investigations of interviewing as a process have sought facts relating to these very issues. There are other questions to be raised about interviewing as a process which do not concern tests. One would like to know, for instance, what factors tend to facilitate free and frank expression, even of personal matters. One would like to know the conditions for gaining rapport, or the factors which lead to blocking, inhibition, tend-

ency to distort, or to prevaricate. If interviewing could be analyzed and studied experimentally in answer to such questions, the use of the interview in counseling and case-work could be placed on a surer basis.

By validity is meant the degree to which the data secured in an interview correspond to the actual facts. For example, the correspondence between what a person says concerning his income and his actual income is a matter of validity. This again may be split into two parts: The accuracy of report of the interviewee; and the accuracy of judgment of the interviewer as to what the interviewee says and his interpretation of it. To study the first of these meanings the common method has been to compare the subject's report with independent sources of evidence. To measure the second it is possible to have a person listen to a recorded interview and then proceed to write his own summary of it and to answer certain questions about it and compare his summary or answers with the summary or answers of what is acknowledged to be the truth or the summary or answers of some acknowledged expert. The validity of the interview may also be influenced by the actual process itself, arising not only in the bias of the subject or in the interviewer's interpretation of what he says, but also in the selection of questions and directions of the interview. Rice[1] gives convincing evidence that the bias of the interviewer colors both the conduct and the results of the interview.

The validity of the interview may also be thought of in terms of the accuracy of judgment of the interviewer as the result of his interviewing experiences. Clark,[2] for instance, has studied the relationship between an interviewer's estimates of a student's ability and his actual class marks.

While considering the validity of an interview it might be added that *accuracy* is not always a desideratum in interviewing. The counselor who is using the interview to understand his client better accepts whatever is said in the interview as an expression of the other's personality. The client may be definitely, although unconsciously, distorting the truth by projecting his own feelings on to others, introjecting other persons' attitudes to himself, or rationalizing his position by unsound arguments. The interviewer, instead of being distressed at these distortions of the truth, accepts them as evidence of the indi-

---

[1] Rice, S. A.: "Contagious Bias in the Interview." *American Journal of Sociology*, Vol. xxxv, November, 1929, pp. 420–423.

[2] Clark, E. L.: "Value of Student Interviews." *Journal of Personnel Research*, Vol. v, 1926, pp. 204–207.

vidual's personal needs. So the criterion of validity in the technical sense does not apply in all kinds of interviews.

The reliability of an interview refers to accuracy more in the sense of the consistency of the interview. To what extent would an interview if repeated yield the same results, right or wrong? This again splits into two questions: (1) To what extent will a subject give the same answers to two different interviewers under different conditions, and (2) to what extent will two interviewers make the same interpretation (without regard to whether right or wrong) of a person's testimony.

The first of these can be studied by having the same persons interviewed independently by two interviewers using the same schedule and comparing the answers. This is somewhat artificial, as it is difficult to make the repetition of an interview seem a reasonable task to a subject. Bingham has also used reliability in the sense of testing that form of the interview which "draws from the largest proportion of the people who know about the event the same or nearly the same answer."[1]

One may test reliability in answer to the second question by having again two independent persons listen to a recorded interview and then write a summary of it or answer questions about it. This time the two summaries or answers would be compared with each other as in contrast with the check on validity in which they were compared with the truth of the report or answers of an expert.

Reliability of the interview may also be thought of in relation to judgments of personality made in the interview situations as well as the actual material which the interview yields. It was in this sense that Hollingworth spoke of testing the reliability of the interview when he correlated the character ratings made by a number of interviewers who had interviewed the same groups of men.[2]

Methods of studying the interviewing process with reference to eliciting freer and frank expression, gaining rapport, or studying the factors leading to blocking, inhibition, tendency to distort, or to prevarication are not at all clear. Indeed, by ordinary observations, it is difficult to recognize the presence of these factors. One might attempt to measure free experience by counting the number of words spoken by the subject, but this would be extremely rough. Counting ideas expressed might come a little closer to what is wanted, but involves

---

[1] Bingham, W. V., and Moore, B. V.: *How to Interview.* Harper & Bros., 1931, 1934, p. 192.

[2] Hollingworth, H. L.: *Vocational Psychology and Character Analysis.* D. Appleton Company, 1929.

more judgment on the part of the experimenter. Any attempt to rate these ideas as to their emotional tone involves still more judgment. Judgment would also play a large part in attempting to evaluate the responses according to the types of adjustment, complexes, and mechanisms they represent. One might attempt to measure blocking by the time interval between question and answer.

All of these suggestions imply a somewhat controlled situation with prepared questions and answers. Interviewing, which is a natural conversation between interviewer and subject, would be still harder to study as a process because by its very free nature comparisons would be difficult to make between process and outcome.

Were it possible to invent measures of the results of interviewing it would be desirable to study the factors which cause variations in these results. These factors may be roughly grouped into four classifications: (1) Factors inherent in the subject, (2) factors residing in the interviewer, (3) factors in the general situation in which the interview is conducted, and (4) factors in the form and content of the interview.

Factors Inherent in the Subject.—The study of individual differences in response in the interview situation is an unbroken field. Anyone who has interviewed recognizes these individual differences immediately—some respond freely, frankly and honestly; others are reserved, inhibited, "cagey." These differen cesmay be related to such factors as

a. Age of subject.
b. Intelligence of subject.
c. Sex of subject.
d. Race of subject.
e. Socio-economic level of subject.
f. Language facility of subject.
g. Emotional need of the subject.
h. Emotional security of subject.
i. Subject's attitude (transference) toward interviewer.
j. Subject's previous acquaintance with interviewer.
k. Purpose with which subject comes to interview.

Factors Inherent in the Interviewer.—Then it should be recognized that the interview is a bi-polar affair. The response is conditioned as much by the interviewer as by the subject. Such factors as the following may condition the interview product.

a. Age of interviewer.
b. Sex of interviewer.

c. Intelligence of interviewer.
d. Race of interviewer.
e. Socio-economic level of interviewer.
f. Position or authority of interviewer with reference to the subject.
g. Personality of the interviewer (social warmth, sympathy, outgoingness, human interest, vitality).
h. Social outlook of the interviewer.
i. Psychological understanding of the interviewer (his ability to sense or feel the purposes, needs or drives, mechanisms of the subject).
j. Interviewer's previous acquaintance with subject.
k. Use interviewer can be to subject.
l. Interest of interviewer to subject.

Factors in the General Situation in Which the Interview Is Conducted.

a. Place of interview.
b. Time of interview.
c. Persons present.
d. First, second, or subsequent interview.
e. Experiences of subject directly preceding the interview.
f. Emergency character of interview (see *g.* under factors inherent in the subject).
g. Directions given subject preceding the interview.
h. Voluntary *vs.* non-voluntary nature of interview.

Factors in the Form and Content of the Interview.

a. Content of questions.
b. Form of questions.
c. Interpretations, suggestions, or other reactions of interviewer.
d. Telling subject purpose of the interview.
e. Relieving subject as to identity of the interviewer.
f. Encouragement given by interviewer.
g. Remarks interpolated by examiner during the interview.

These lists are not intended to be complete and they can probably be extended by anyone who has had interviewing experience.

Such is a theoretical analysis of the study of the interviewing process. Actually research on interviewing is beset by extraordinary difficulties. These difficulties are of two kinds. One concerns the question of standardization. Research which involves rigid comparisons of the results suggests the desirability of much uniformity and control in the interviewing. One would like to know how persons respond when one factor is varied and the interviewer, the interviewee, the conditions of the interview, and the form of the interview are kept constant. This implies that the interview should proceed by the use of prepared questions. But the use of prepared questions defeats the

nature of the interview process which is a free exchange of opinion between interviewer and subject.[1] Again it is very difficult to arrange a standard interview procedure for purely experimental purposes which will, at the same time, be a real situation, for an interview, above everything else, must be in the stream of living. The interviewer

[1] In planning an investigation of the interview as a process I drew up a list of questions which I planned to ask and submitted them to a number of psychiatrists, mental hygienists, and child guidance experts. Their replies are most illuminating and indicate the difficulties which beset one who wishes to apply rigid tests to the interviewing process.

Jessie Taft wrote: "In the type of situation which you describe, it seems to me that all you can hope to discover is how various children react to being asked questions unrelated to their purposes, questions which they have not sought, but which they cannot avoid. In my opinion, almost any question is related to attack; at any rate, it tends to bring out fear, defense, or resistance even on the part of an adult who expects and permits it. Also, I have found in my contacts with people that one learns little from any interview unless it goes beyond the question stage to some form of spontaneous expression. The therapeutic interview depends largely on the fact that the person comes for it."

Bruce B. Robinson wrote: "There seems to be too little attention to the important fact that an interview is a matter of the relationship established between the interviewer and the subject, and that the conditions which determine this relationship can never be made constant. . . . I feel that we cannot get the improvement in interviewing which we all desire until the questionnaire method has been abandoned and an informal conversational type developed."

Phyllis Blanchard wrote: "I would suspect that certain of your questions, at least with some children, would elicit a conventional or defensive reply; that the reply will be what the child thinks he ought to feel or what he ought to say, or an effort to conceal certain feelings from himself. . . . It seems to me that the real feelings of a child may have come out rather in a relationship in which they are well acquainted with someone and have some security with the person than in reply to formal questions."

Karl A. Menninger wrote: "One of the important advantages of the interview over the questionnaire or the standardized test is its flexibility, and this flexibility is gained by avoiding the stereotypy of the test situation. An attempt to rigidly control as many factors as possible in an interview robs it of its greatest value. . . . Direct questioning of children, such as you propose to do in this experiment, can be expected to bring out replies on a superficial basis only, unless previous interviews have already established rapport. If the questions are given according to a fixed routine, it would not be possible for an interviewer to follow up a promising lead furnished by the child, and hence one would not be able to discover individual differences."

Augusta F. Bronner wrote: "In my own experience I find it necessary first of all to establish friendly relations through trying to discover why the child thinks he is being questioned and trying to make clear to him the purpose of the interview itself."

must be his real self, the subject of the interview must be real and meaningful to both parties. Something is vitally wrong with the interview which is hatched up for the purposes of studying the process.

In the second place, there is the difficulty of recording the interview. A record which the interviewer makes following the interview is imperfect and inaccurate. It might be possible to have a stenographer hidden behind a screen, but there is always the risk of detection, and the possibility that some of the conversation will be lost when voices are low.

In recent years important studies have been made in the improvement of automatic electrical recording. The goal is faithful recording by a pick-up of the voices in the natural interviewing situation. The dictaphone gives relatively faithful recording but is inadequate for the purpose of recording interviews in that one must speak directly into the mouthpiece as into a telephone and the record is transcribed onto perishable wax.

Modern electrical recording uses apparatus which consists of three parts. First, there is the sensitive crystal microphone which can pick up extremely low sounds and transmit them with fidelity in both the lower and upper vibration frequencies. Such a microphone will transmit the ordinary voice which may be on the opposite side of the room. The second part of the apparatus is the amplifying system. The third part is the recording mechanism. This consists of a turntable on which discs (phonograph records) revolve. A special cutting head will cut the record of the sound on to the grooves in the disc. Less expensive are the aluminum discs in which the record is pressed rather than cut. More expensive and also more satisfactory is the acetate (black) disc in which the record is actually cut out giving the more faithful reproduction with less scratching. For playing back the record the "speaker" is a fourth feature of the apparatus.

The apparatus is now fairly well perfected. The difficulty comes in its use. In the first place there is the problem of the concealment of the microphone. This may be accomplished by placing it in a lamp base or screwing it into the socket of a lamp or hiding it in what looks like a bookend. But any fabric placed over a microphone tends to dampen and distort the sound to some extent. Probably it is not necessary to conceal the microphone in most circumstances. It is the last thing to enter the mind of most subjects that what they are saying in any interview is being recorded, and the microphone will be considered by them as part of the scientific background of the room.

More serious are the acoustics of the room. A sensitive microphone will pick up all sounds without discrimination. Until one has worked with automatic recording apparatus he is unaware of the relative intensity of sounds and of the great amount of selection which we ordinarily make in listening to the human voice with a background of even a greater volume of sound of passing automobiles, footfalls, and the like. The microphone will record not only the interview but the nearby typewriter, the heels tapping their way down the corridor, the high-pitched voices of the class at recess in the nearby playground (even though the windows be closed). Good recording demands something approximating sound-proof conditions.

Then one must also consider the walls and windows of the room in which the interviewing is conducted. These serve as reflectors and throw the sound back to be repeated in the microphone to give it a definite blur. In listening to a person we select his first words and ignore the reflected sounds. It is very difficult to make this selection from recorded speech. This reflection may be deadened by suitable wall-hangings.

A number of commercial companies[1] now assemble electrical sound recording apparatus which is reasonably perfect from the mechanical point of view. Sound recording apparatus is also assembled from standard equipment by individuals.

Research on the interviewing process involves peculiar difficulties which are not found in testing. By its very nature interviewing must be relatively free and unstandardized, and yet the research requires a certain amount of standardization in order to eliminate the influence of concomitant variables. Research on the interviewing process is probably best carried out by a person whose work is interviewing and who must interview day after day under relatively uniform conditions. It is hoped that this paper will come to the attention of such persons and that they will be stimulated to undertake studies of the interviewing process.

---

[1] The Presto Recording Corporation, 139 West 19th Street, New York City; Fairchild Aerial Camera Corporation, 62-10 Woodside Avenue, Woodside, N. Y.; the Speak-O-Phone Recording and Equipment Company, 23 West 60th Street, New York City; the Universal Microphone Company, 424 Warren Lane, Inglewood, California, and the Sound Apparatus Company, 150 West 46th Street, New York City.

ACCURACY *VERSUS* ERROR AS A CRITERION IN CHILDREN'S SPEECH*

EDITH A. DAVIS

It is now thirty years since the pioneer study of oral errors in the speech of public-school children was undertaken by G. M. Wilson in Connersville, Indiana.[3] Investigators since that time have agreed that intensive drives for accuracy have never succeeded in eradicating from the speech of a given group of children even the most "heinous" errors, and the general tone of the findings reported in the literature is extremely pessimistic. Nearly all these studies, however, are somewhat unfair to the children involved, since teachers merely record the errors as they hear them, but pay no attention to the occurrence of the *correct* forms of the constructions. An exact transcription of actual speech is, therefore, essential, because there is evidence indicating that the error compels attention while the correct form passes unnoticed. (Menner,[2] recording the speech of forty adults, found it much more difficult to "hear" correct verb forms than incorrect ones.)

In her analysis of fifty remarks obtained from each of four hundred thirty-six children ranging in age from five and one-half to nine and one-half years, the present writer[1] listed the various types of error in order of frequency and computed the ratio of each type to each thousand words recorded, also its percentage ratio to all the errors recorded. Errors were slightly more frequent in boys than in girls and in children from the lower occupational groups than in those from the upper ones, and the ratio of errors to total speech decreased slightly with advancing age.

Such an analysis, however, does not tell us the extent to which the older child, as compared with the younger, has achieved mastery over a specific construction, nor whether a given type of error occurs almost every time that particular construction is used or only once or twice in a hundred chances. The important thing is not that once in a while a child says "more safer" or "goodest," but how often he compares an adjective correctly. We all know that children, and adults as well, often use "lay" for "lie," but we need to know whether they ever use "lie" correctly. "Mans" and "sheeps" offend the ear, while "men" and "sheep" may easily pass unnoticed. The writer has just completed an analysis with the object of determining

* From the Institute of Child Welfare, University of Minnesota.

just such facts. The findings show general mastery of inflections and improvement with age to greater extent than we should expect from merely scanning the literature on errors.

This analysis, which undertook to compare the number of times a given construction was used correctly with the number of times it was used erroneously, offered some problems of method. Not all errors can be studied in this way, for the mathematics involved would be beyond human reckoning. Thus, there were forty-four instances of wrong word order, obviously of very slight importance in a mass of material approximating one hundred twenty thousand words, yet the possible arrangements of these words are practically incalculable. Similarly, the wrong preposition or conjunction was used one hundred fourteen times, but it would have been exceedingly difficult to determine the mathematical chances for inaccuracy in the use of these words.

The constructions finally selected for study were those most often used erroneously, providing error or accuracy could be clearly determined. The types of construction and the method of determining accurate use were as follows:

(1) Agreement of verb and subject. This construction was counted only when an inflected form was used, not if the singular and plural or first and third person were alike. Invariable forms of verb or pronoun (could, can, did, sent, who) offer no opportunity for agreement or disagreement.

(2) Correct formation of irregular preterit or past participle.

(3) Use of "isn't," "aren't," "hasn't," or "haven't," instead of "ain't," or "hain't."

(4) Use of simple negative. This construction presented some difficulties. Theoretically, any sentence containing one negative might include a superfluous one. Actually, since "There isn't not any more" was extremely infrequent, the opportunity for double negatives was considered to be limited to such sentences as "I never knew that," "This one hasn't any," or "He can hardly stand up." If all negative sentences had been included, the percentage of accuracy even for the youngest children would have been very high, whereas the method finally employed showed definite improvement with age.

(5) Use of "lie" distinguished from "lay."

(6) Avoidance of redundant "here" or "there." "This here" and "that there" were very common. Another redundancy frequently encountered was the interpolation of a pronoun between

subject and verb ("This man, he"); but, since this error might occur in every clause, this particular construction was not tabulated. As with the negative constructions, its inclusion would have made the total percentage of accuracy for redundant expressions exceedingly high.

(7) Use of "an," not "a," before words beginning with a vowel sound.

(8) Correct use of adverb, not adjective. ("I don't like it very good.")

(9) Correct formation of irregular plural nouns.

(10) Correct comparison of adjectives.

(11) Correct inflection of pronouns. The forms tabulated were variants from "I," "he," "it," and "you," which were taken as standards. Thus, all correctly inflected feminine and plural forms of the personal pronouns, and possessive and objective case forms, were listed. "Whose," "these," and "those" were also counted as inflected forms. If every correct usage of "I," "he," "it," and "you" had been included, the percentage of accuracy for pronoun forms would have been practically one hundred per cent.

FINDINGS

Several different ratios and indices were computed for each type of error studied, for two reasons. In the first place, since the findings of a number of investigations have been published only in part, it is hoped that readers having in their possession material tabulated according to one or another of these plans will be able to select the measure suited for comparison with their own data. Secondly, it is very important to ascertain whether the easy method of relating the number of errors to the total number of words used by a given group of children gives the same final picture as the much more laborious one of relating incorrect and correct usage. Table I summarizes these numerical expressions for all the subjects used in this study.

The first two columns in Table 1 show the mean number of times that each child used the various constructions correctly and the number of times he used them incorrectly. The computations in the other columns were made from total occurrences, so that the resulting ratios are slightly different from those that would be obtained from dividing the first column by the second column. The difference is very slight, however, except in the case of constructions seldom used. The findings indicate, in the case of agreement between verb and

subject, that the construction was used correctly 26.3 times per child, and erroneously 1.8 times per child. There were one hundred one instances of correct usage in each thousand words recorded, and seven instances of erroneous usage. There was approximately one correct usage to each ten words recorded, and one error to each one hundred forty-two words. The ratio of times correct to times incorrect is fourteen to one. Stated in terms of percentage, the use of this construction was correct in 93.5 per cent of possible cases. There were sixty-five errors out of one thousand chances to use the construction.

TABLE I.—ANALYSIS OF CORRECT AND INCORRECT USAGE OF VARIOUS GRAMMATICAL CONSTRUCTIONS BY FOUR HUNDRED THIRTY-SIX CHILDREN

Type of construction	Mean occurrence per child		Number occurrences per thousand words		Ratio to total words used		Ratio of times correct to times incorrect	Percentage of correct use	Errors per thousand chances
	Correct	Incorrect	Correct	Incorrect	Correct	Incorrect			
Agreement of verb and subject	26.3	1.8	101.1	7.0	1:10	1:142	14:1	93.5	65
Inflection of pronouns	16.5	0.3	63.4	1.1	1:16	1:873	55:1	98.2	18
Redundancy	11.4	1.1	43.6	4.2	1:23	1:237	10:1	91.2	88
Irregular tenses	6.9	0.2	26.7	0.9	1:37	1:1102	29:1	96.7	33
Irregular nouns	0.8	0.2	3.1	1.0	1:325	1:1022	3:1	75.9	241
Use of "ain't"	0.9	0.3	3.1	1.2	1:319	1:805	3:1	71.6	284
Use of negative	0.7	0.2	2.6	0.7	1:387	1:1493	4:1	79.4	206
Use of "an"	0.7	0.3	2.6	1.4	1:380	1:732	2:1	65.8	341
Comparison of adjectives	0.6	0.02	2.4	0.06	1:407	1:16,211	40:1	97.5	24
Adverb, not adjective	0.06	0.23	0.24	1.0	1:4203	1:1041	0.2:1	19.8	801
Use of "lie"	0.01	0.26	0.06	1.0	1:16,211	1:1004	0.06:1	5.8	942

The necessity for considering incidence as well as accuracy appears when we compare the inflection of pronouns with the comparison of adjectives. These two types of construction rank at the top of the list for accuracy, yet a pronoun is inflected correctly once for every sixteen words and incorrectly once for every eight hundred seventy-three words. The corresponding figures for the adjectives are four hundred seven and sixteen thousand two hundred eleven. The opportunity to inflect a pronoun occurred approximately seventeen times for each child, while adjectives were compared less than once.

All the measures used revealed definite improvement with advancing age in the mastery of these difficult constructions, except in the use of redundant words. In general, the number of correct usages increased and the number of incorrect usages decreased with advancing

TABLE II.—MEAN NUMBER OF CORRECT AND INCORRECT USAGES BY EACH CHILD

Type of construction	Five and one-half years		Six and one-half years		Nine and one-half years		All ages	
	124 boys	124 girls	32 boys	31 girls	62 boys	63 girls	218 boys	218 girls
Agreement of verb and subject.								
Correct	24.7	25.0	28.5	30.7	26.1	29.1	25.6	27.0
Incorrect	1.9	1.9	1.9	2.0	1.9	1.3	1.9	1.7
Inflection of pronouns.								
Correct	11.3	13.3	16.9	15.6	23.8	26.1	15.7	17.3
Incorrect	0.3	0.4	0.3	0.1	0.2	0.2	0.3	0.3
Redundancy.								
Correct	12.4	10.4	15.0	11.5	9.7	10.9	12.0	10.7
Incorrect	1.1	0.6	0.5	0.6	1.5	2.0	1.2	1.0
Irregular tenses.								
Correct	3.7	4.2	6.1	7.3	12.6	13.4	6.6	7.3
Incorrect	0.3	0.2	0.2	0.3	0.3	0.1	0.3	0.2
Irregular nouns.								
Correct	0.2	0.2	0.2	0.4	2.3	2.2	0.8	0.8
Incorrect	0.2	0.3	0.2	0.3	0.3	0.3	0.2	0.3
Use of "ain't."								
Correct	0.7	1.0	1.0	1.2	0.6	0.5	0.7	0.9
Incorrect	0.5	0.4	0.5	0.3	0.2	0.03	0.4	0.2
Use of negative.								
Correct	0.6	0.6	0.6	1.0	0.8	0.7	0.7	0.6
Incorrect	0.2	0.2	0.2	0.3	0.2	0.1	0.2	0.1
Use of "an."								
Correct	0.5	0.3	0.5	0.7	1.6	0.9	0.8	0.6
Incorrect	0.6	0.3	0.2	0.3	0.2	0.1	0.4	0.3
Comparison of adjectives.								
Correct	0.3	0.5	0.8	0.9	1.0	0.9	0.6	0.7
Incorrect	0.02	0.01	0.03	0.00	0.00	0.05	0.01	0.02
Adverb, not adjective.								
Correct	0.02	0.02	0.03	0.1	0.3	0.3	0.04	0.08
Incorrect	0.19	0.29	0.31	0.13	0.29	0.27	0.19	0.26
Use of "lie."								
Correct	0.02	0.02	0.03	0.00	0.00	0.00	0.02	0.01
Incorrect	0.27	0.33	0.12	0.16	0.19	0.27	0.23	0.29

age, both absolutely and when related to the total number of words used. Girls tended to be slightly superior to boys both in using more words correctly and in making fewer errors. These differences are

TABLE III.—NUMBER OF CORRECT AND INCORRECT USAGES PER THOUSAND WORDS

Type of construction	Five and one-half years		Six and one-half years		Nine and one-half years		All ages	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
Agreement of verb and subject.								
Correct	109.6	108.1	112.2	116.2	82.7	87.6	100.0	102.0
Incorrect	8.6	8.1	7.5	7.7	6.1	4.0	7.6	6.5
Inflection of pronouns.								
Correct	50.3	57.6	66.7	59.0	75.3	78.5	61.0	65.0
Incorrect	1.5	1.7	1.2	0.5	0.7	0.5	1.2	1.1
Redundency.								
Correct	55.1	45.1	59.0	43.5	30.5	32.7	47.0	40.0
Incorrect	4.9	2.8	2.1	2.2	4.9	6.1	4.5	3.9
Irregular tenses.								
Correct	13.1	18.1	23.9	27.7	39.8	40.4	26.0	27.0
Incorrect	1.5	0.9	0.6	1.1	0.8	0.2	1.1	0.7
Irregular nouns.								
Correct	0.8	0.9	0.9	1.7	7.2	6.6	3.1	3.0
Incorrect	0.7	1.3	0.7	1.0	0.9	0.9	0.8	1.1
Use of "ain't."								
Correct	3.1	4.4	4.1	4.6	2.0	1.4	2.9	3.4
Incorrect	2.1	1.6	1.8	1.1	0.6	0.1	1.5	1.0
Use of negative.								
Correct	2.9	2.4	2.5	3.7	2.5	2.0	2.7	2.5
Incorrect	0.8	0.8	0.9	1.0	0.5	0.2	0.7	0.6
Use of "an."								
Correct	2.0	1.5	2.0	2.6	5.2	2.8	3.1	2.1
Incorrect	2.6	1.5	1.0	1.2	0.6	0.4	1.7	1.0
Comparison of adjectives.								
Correct	1.6	2.1	3.2	3.4	3.3	2.6	2.4	2.5
Incorrect	.07	0.03	0.12	0.00	0.00	0.14	0.05	0.07
Adverb, not adjective.								
Correct	0.1	0.1	0.1	0.5	0.3	0.5	0.18	0.29
Incorrect	0.9	1.2	1.2	0.5	0.9	0.8	0.9	1.0
Use of "lie."								
Correct	0.1	0.1	0.1	0.00	0.00	0.00	0.07	0.05
Incorrect	1.2	1.4	0.5	0.6	0.6	0.8	0.9	1.09

shown in Tables II, III, IV, V, and VI. In the case of several of the constructions studied, tabulation of errors alone would not have revealed the changes that take place with age. The incidence of

TABLE IV.—RATIO OF CORRECT TO INCORRECT USAGES*

Type of construction	Five and one-half years		Six and one-half years		Nine and one-half years		All ages	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
Agreement of verb and subject	13.0	13.0	15.0	15.0	14.0	22.0	13.0	15.0
Inflection of pronouns.	33.0	33.0	54.0	121.0	114.0	164.0	52.0	59.0
Redundancy	11.0	16.0	28.0	20.0	6.0	5.0	10.0	10.0
Irregular tenses	11.0	19.0	38.0	25.0	49.0	169.0	23.0	39.0
Irregular nouns	1.5	0.7	1.2	1.7	7.9	6.9	3.3	2.7
Use of "ain't"	1.5	2.8	2.2	4.2	3.6	15.0	1.9	3.5
Use of negative	3.5	2.9	2.9	3.7	5.0	10.7	3.7	4.0
Use of "an"	0.8	1.0	2.0	2.1	8.5	7.4	1.9	2.0
Comparison of adjectives	22.0	61.0	26.0	.....	.....	18.0	45.0	38.0
Adverb, not adjective.	0.1	0.1	0.1	1.0	0.3	0.6	0.2	0.3
Use of "lie"	0.1	0.1	0.2	.....	.....	.....	0.1	0.05

* Times correct: Times incorrect :: x:1.

TABLE V.—NUMBER OF ERRORS PER THOUSAND CHANCES

Type of construction	Five and one-half years		Six and one-half years		Nine and one-half years		All ages	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
Agreement of verb and subject	73	70	63	62	68	43	70	60
Inflection of pronouns.	30	29	18	8	9	6	19	17
Redundancy	82	60	34	48	138	158	88	89
Irregular tenses	81	49	25	38	20	6	41	25
Irregular nouns	477	603	461	364	112	127	207	272
Use of "ain't"	401	262	312	191	216	62	345	223
Use of negative	223	255	259	210	167	85	210	201
Use of "an"	565	494	333	323	105	119	349	330
Comparison of adjectives	43	16	37	0	0	51	22	27
Adjective used for adverb	889	923	909	500	750	630	839	770
Use of "lie"	919	932	800	1000	1000	1000	926	954

TABLE VI.—RATIO OF CORRECT AND INCORRECT USAGES TO TOTAL WORDS USED*

Type of construction	Five and one-half years		Six and one-half years		Nine and one-half years		All ages	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
Agreement of verb and subject.								
Correct	9	9	9	9	12	11	10	10
Incorrect	116	124	133	130	165	252	132	153
Inflection of pronouns.								
Correct	19	17	15	17	13	13	16	15
Incorrect	649	574	812	2048	1508	2094	843	904
Redundancy.								
Correct	18	21	17	23	33	30	21	25
Incorrect	202	354	478	455	204	162	222	254
Irregular tenses.								
Correct	60	55	42	36	25	25	39	36
Incorrect	681	1063	1624	910	1225	4189	897	1899
Irregular nouns.								
Correct	1213	1148	1160	585	138	152	323	327
Incorrect	1329	755	1353	1024	1089	1047	1236	876
Use of "ain't."								
Correct	317	224	246	215	490	698	345	297
Incorrect	473	638	541	910	1782	10,472	654	1033
Use of negative.								
Correct	349	411	406	273	392	487	371	403
Incorrect	1213	1196	1160	1024	1960	5236	1391	1607
Use of "an."								
Correct	490	652	507	390	173	355	318	466
Incorrect	377	668	1015	819	1634	2618	592	948
Comparison of adjectives.								
Correct	634	471	312	292	302	381	412	402
Incorrect	13,954	28,711	8121	0	0	6981	18,545	14,461
Adverb, not adjective.								
Correct	9303	9570	8121	2048	3267	2094	5563	3403
Incorrect	1163	797	812	2048	1089	1232	1070	1015
Use of "lie."								
Correct	9303	9570	8121	0	0	0	13,908	19,282
Incorrect	821	700	2030	1638	1634	1232	1113	918

* In reading the table, the figure 1 with the ratio sign should be supplied before each figure given in the table. That is, both boys and girls aged five and one-half years made a verb and subject agree once for every nine words used. Boys aged five and one-half used a verb and subject in disagreement once for every one hundred sixteen words used, and girls aged five and one-half once for every one hundred twenty-four words used.

errors in the formation of irregular plural nouns is relatively unchanged from one age level to the next, yet the correct usage increases enormously. On the other hand, older children, especially girls, say "ain't" less often than younger children, but it is also true that they use "isn't" and "aren't" less often than younger children do.

The ratio of correct to incorrect usages, shown in Table IV, indicates that the amount of improvement with age is greater than has been supposed. This is, of course, due to the fact that correct usage of most constructions increases at the same time that incorrect usage decreases. An exception to this rule is agreement of verb and subject, which is less per thousand words at nine and one-half than at five and one-half years.

Since young children learn language from those about them, we should expect to find greater accuracy of speech in children coming from superior homes than in those coming from poorer homes. Table VII, which shows the degree of mastery of the constructions studied, indicates that for some of these constructions there was a decided difference between children coming from the upper and those coming from the lower occupational groups. Educators may well be gratified to learn that it was among children from the less privileged homes that the greatest improvement took place during the early school years. For instance, in the elimination of "ain't" and of the use of the double negative children of nine and one-half years from the upper occupational groups were not greatly superior to children of five and one-half years, while children from the lower occupational groups showed decided improvement. Since these changes were almost certainly due to the cultural advantages of the school system (certainly not to the radio, the movies, or the comic strips), the fact should afford great encouragement to educators who may have felt that they waged a losing fight against these very errors. Table VII shows some of these changes with age.

While these findings are fairly convincing for the constructions used with great frequency, the findings for constructions of rare occurrence must be interpreted with extreme caution. Thus, although a girl of nine and one-half years made a mistake in the inflection of pronouns only six times out of one thousand chances, whereas she said "lay" instead of "lie" one thousand times out of one thousand chances, we must remember that in the language samples analyzed a girl of this age inflected, on the average, twenty-six pronouns, while only about one girl in three happened to use the word "lie" at all.

TABLE VII.—PERCENTAGE OF ACCURACY FOR VARIOUS TYPES OF CONSTRUCTION BY AGE, SEX, AND OCCUPATIONAL GROUPS

Type of construction	Five and one-half years		Six and one-half years		Nine and one-half years	
	Boys	Girls	Boys	Girls	Boys	Girls
Agreement of verb and subject.						
Upper	93.5	92.5	95.3	94.9	93.5	99.2
Lower	91.6	93.6	92.3	92.4	92.7	99.0
All	92.5	93.0	93.7	93.8	93.2	99.1
Inflection of pronouns.						
Upper	96.7	96.7	93.3	99.2	96.8	99.4
Lower	97.3	97.4	98.0	99.1	94.6	99.4
All	97.0	97.1	98.2	99.2	95.7	99.4
Redundancy.						
Upper	94.0	95.0	98.3	96.8	89.4	82.1
Lower	89.3	93.5	94.2	93.0	83.3	86.2
All	91.8	94.1	96.6	95.2	86.2	84.2
Irregular tenses.						
Upper	94.4	95.6	99.1	96.0	98.0	99.5
Lower	89.4	94.5	95.1	96.7	98.0	99.3
All	91.7	95.1	97.5	96.2	98.0	99.4
Irregular nouns.						
Upper	70.8	39.0	44.4	60.0	91.1	82.2
Lower	30.0	40.9	75.0	64.7	85.7	89.5
All	52.3	39.7	53.8	63.6	88.7	86.5
Use of "ain't."						
Upper	85.3	82.7	65.4	75.7	88.6	94.1
Lower	38.0	60.3	72.7	92.8	56.2	93.3
All	59.9	73.8	68.7	80.8	78.4	93.7
Use of negative.						
Upper	89.6	78.0	70.0	95.4	81.8	95.0
Lower	62.2	68.6	76.5	56.2	85.2	88.9
All	77.7	74.5	74.1	78.9	83.3	91.5
Use of "an."						
Upper	48.6	54.3	33.3	75.0	96.1	90.9
Lower	37.3	46.3	86.7	60.0	84.1	85.3
All	43.5	50.6	66.7	67.7	89.5	88.0
Comparison of adjectives.						
Upper	92.3	100.0	93.7	100.0	100.0	95.8
Lower	100.0	96.1	100.0	100.0	100.0	94.1
All	95.6	98.4	96.3	100.0	100.0	94.8
Adverb, not adjective.						
Upper	10.0	4.5	14.3	42.8	40.0	25.0
Lower	11.8	11.8	0.0	100.0	0.0	71.4
All	11.1	7.7	10.0	50.0	25.0	37.0
Use of "lie."						
Upper	10.0	15.4	0.0	0.0	0.0	0.0
Lower	5.9	3.2	25.0	0.0	0.0	0.0
All	8.1	6.8	20.0	0.0	0.0	0.0

We are not prepared to say in how many instances a cleverly worded question would have elicited a correct distinction between "lay" and "lie," but we do know that in the seventeen instances of spontaneous usage by girls of this age, every one was wrong. On the other hand, by the age of nine and one-half years girls appeared to have achieved virtual mastery of the three most commonly used constructions studied; namely, agreement of verb and subject, the irregular formation of preterit and past participle, and the inflection of pronouns, since their usage was correct in from ninety-nine to ninety-nine and five-tenth per cent of instances. This was true for girls from the lower as well as for those from the upper occupational groups. The percentage of accuracy for boys of nine and one-half years varied for the same constructions from ninety-two and seven-tenths to ninety-eight per cent.

In a way, the performance of the younger children was even more amazing, when we consider that the errors noted are among the most arresting ones in children's speech. For only one group of subjects (five and one-half-year-old boys from the lower occupational groups) was the percentage of accuracy for any of these frequently used constructions lower than ninety per cent.

There seems to be no escaping the conclusion that reports based on frequency of error alone are inadequate and misleading. One should know the frequency of occurrence of a given construction in a given sample of language before attempting to evaluate the errors in children's speech. It is easy to become so preoccupied with the one hundred forty-one times a group of children say "ain't" as to lose sight of the three hundred fifty-six times they say "isn't."

REFERENCES

1. Davis, Edith A.: *The Development of Linguistic Skill in Twins, Singletons with Siblings, and Only Children from Age Five to Ten Years.* No. 14, Institute of Child Welfare Monograph Series, Univ. of Minnesota Press, Minneapolis, 1937.
2. Menner, R. J.: "The Verbs of the Vulgate." *American Speech*, Vol. 1, January, 1926, pp. 231–240.
3. Wilson, G. M.: Errors in language of grade pupils." *Educator-Journal*, December, 1909.

COMPARISON OF EARLY BINET RECORDS WITH COLLEGE APTITUDE TEST SCORES

GERTRUDE HILDRETH

Lincoln School, Teachers College, Columbia University

The extent to which rank in mental ability or scholastic aptitude as measured by standardized tests remains constant or varies as individuals increase in age is of considerable interest to psychologists and educators. Several research studies relating to this problem have previously been published.* The reports are of two types; first, those that study mental status as measured repeatedly with the same instrument, and, second, those that employ different instruments at varying age levels. Because of the inadequacy of the Stanford-Binet 1916 Revision for the measurement of older subjects, most studies, particularly those involving gifted children, have employed more adequate group intelligence tests as an alternative measure.

This report is concerned with the mental status of pupils determined during the elementary-school period with the 1916 Revision of the Stanford-Binet Test, and a later measure of intellectual aptitude made during the senior high-school year with the American Council on Education Psychological Examination. The data were derived from the records of all pupils attending a private school for whom these two test records were available at the time the study was undertaken. There were seventy such cases in all, thirty-eight girls and thirty-two boys. Table I shows the distribution of the group in chronological age at the time of the Binet test as well as the distribution of IQ's obtained. The median chronological age was nine years, nine months, and all of the children were under thirteen years of age at the time of the first test. The median Stanford Binet IQ was 120. At the time of the second test pupils were between sixteen and eighteen years of age. These senior-year records were obtained over a three-year period from three successive classes in the high school. The earlier Binet examinations were administered by several examiners, all competent and well-trained. The American Council tests were administered by the school psychologist at the beginning of the senior year.

Since scores for the American Council test are not seriously affected by slight differences in age above sixteen years, and a narrow age range

* See bibliography at the end of this article.

characterized the subjects, the scores were uncorrected for age. The effect of age variability on the Binet tests was eliminated by using the IQ as the measure of mental ability. Raw scores for the ACE tests for the years 1935, 1936, and 1937 were made comparable by the use of tables supplied with the manual of directions.

TABLE I.—AGE AND IQ DISTRIBUTION AT THE TIME OF THE FIRST TEST

Chronological age	Frequency	Stanford Binet IQ	Frequency
14-0	1	170	
13-6	0	165	1
13-0	0	160	1
12-6	8	155	2
12-0	4	150	1
11-6	2	145	5
11-0	6	140	2
10-6	4	135	5
10-0	6	130	5
9-6	9	125	5
9-0	7	120	7
8-6	6	115	10
8-0	10	110	14
7-6	3	105	7
7-0	1	100	2
6-6	1	95	3
6-0	1		
5-6	1	Median = 120	
Median = 9-9			

The two sets of scores (See Table II) were made comparable by converting both the Binet and ACE scores into percentile ranks. The Binet IQ's were converted into percentile ranks on the basis of data from an unselected sampling of Binet records for the entire elementary-school population in which these pupils were enrolled. Three hundred five records were used in computing a percentile rank conversion table. The median IQ for the three hundred five cases was 120, the same as for the sample of seventy in this study. Comparable percentile rank data for the ACE scores were obtained by making a percentile score table from the distribution of all senior-year classes for the years 1935, 1936, and 1937. There were one hundred forty-five cases in this distribution.

In comparing the two sets of percentiles, the only question that would arise is whether the one hundred forty-five twelfth-graders

TABLE II.—TEST DATA FOR THE THIRTY-TWO BOYS AND THIRTY-EIGHT GIRLS

Sex	Stanford-Binet examination 1916 revision			American Council Psychological Examination 1935, 1936 and 1937 editions		Gain or loss in percentile rank
	CA	IQ	Percentile	Score	Percentile	
M	8-2	98	2	199	39	37
F	9-4	119	46	198	38	− 6
M	12-4	128	71	236	72	1
M	5-7	163	99	355	100	1
M	10-9	118	44	242	77	33
F	9-6	106	8	173	23	15
M	10-2	152	98	314	97	− 1
F	8-9	116	36	181	28	− 8
F	8-7	111	19	228	67	48
M	12-5	103	4	159	14	10
M	9-5	112	24	222	62	38
M	8-2	110	16	211	49	33
M	12-8	139	91	271	89	− 2
F	8-11	120	48	274	90	42
F	6-0	117	40	233	70	30
M	10-10	129	75	209	48	−27
M	9-11	114	30	187	30	0
F	11-9	108	12	196	37	25
F	9-7	121	51	222	61	10
F	10-4	123	58	184	29	−29
F	12-10	117	40	253	82	42
F	8-5	105	7	164	17	10
F	9-1	99	2	129	6	4
M	11-1	142	93	296	95	2
F	11-5	147	96	286	93	− 3
F	9-11	121	51	153	15	−36
F	8-0	127	70	260	85	15
F	6-7	120	48	188	36	12
M	12-9	116	36	190	37	1
F	8-2	137	88	253	82	− 6
M	8-0	112	24	231	70	46
M	9-6	113	26	158	19	− 7
M	7-8	121	51	189	36	−15
M	8-6	147	96	240	77	−19
M	8-9	137	88	195	41	−47
F	8-6	109	15	168	24	9
F	10-4	131	78	217	59	−19
F	12-5	146	95	238	76	−19
F	9-1	128	71	227	68	− 3
M	9-8	113	26	170	25	− 1
F	9-10	131	78	227	68	−10
F	12-7	139	91	286	93	2
F	12-10	134	85	275	90	5
F	8-0	109	15	102	2	−13
M	11-1	118	44	233	73	29
M	8-3	145	95	290	94	− 1
M	10-11	156	99	326	98	− 1
F	7-5	118	44	189	36	− 8
F	11-0	110	16	153	16	0
M	11-10	104	5	209	48	43
F	10-7	139	91	242	78	−13
F	9-1	125	64	239	76	12
M	10-3	166	99	293	94	− 5
F	8-2	113	26	199	43	17
F	9-2	130	76	176	28	−48
F	11-1	156	99	249	81	−18
F	9-8	147	96	312	96	0
F	12-1	106	8	179	29	21
M	7-6	118	44	227	68	24
F	10-1	115	32	207	49	17
M	12-10	133	82	268	88	6
M	11-1	112	24	239	77	53
M	8-3	112	24	169	25	1
M	9-11	120	48	211	52	4
F	12-11	101	3	201	44	41
F	10-5	108	12	208	51	39
M	12-11	140	92	272	89	− 3
M	7-7	114	30	200	44	14
F	11-7	97	2	131	7	5
M	9-4	113	26	228	69	43

constitute as representative a group for the private school in which they were enrolled as the cross-section of the population in the same school tested under the age of twelve with the Stanford Binet. We have every reason to believe that the groups are entirely comparable, since earlier studies of the same school population indicate that the population of the school is stable, so far as ability and economic status are concerned, from kindergarten to high school. The selective factors operating in the high school are about the same as those operating in the lower grades except that the tuition rate is somewhat higher.

After each pupil's Binet IQ and ACE score were converted into percentile ranks, the Binet percentile rank was subtracted from the ACE rank and a distribution was made of these losses and gains in rank.

Table III shows the distribution of the deviations in percentile ranks for all cases, and for boys and girls separately. Inspection of the distributions for boys and girls separately indicates no reliable difference in the nature of the percentile rank changes for the two sexes.

The following tabulations indicate in general the comparison of initial and final mental status as measured by the two tests: Of those pupils who were at or above the ninetieth percentile on the first test and remained at that percentile on the second test, the number of cases was nine, or sixty-seven per cent. Of those who were ninetieth or better on the first test and remained at the seventy-fifth or higher percentile on the second test, the number was fifteen, or one hundred per cent.

Originally six girls and nine boys rated at the ninetieth percentile or higher on the Binet test. Later, in ACE test scores, only two girls but seven of the boys remained at or above the ninetieth percentile. Of those who were at or below the tenth percentile on the first test and remained at that level, there were two cases or twenty per cent of all those originally testing below the tenth percentile. Of those who were at the thirty-third percentile on the first test or the sixty-seventh percentile on the first test, and remained at those respective points on the second test, the percentages were forty-eight for the first group, thirteen cases; eighty-four for the second group, twenty-one cases.

It is apparent that even in a gifted school population, the relatively more gifted pupils tend to retain their early high rank after a period of several years. There is less tendency for the lowest ranking pupils to remain at the same rank after a period of several years. If a pupil rates high originally he is apt to retain high status (apart from coaching

or other spurious factors). If a child rates low originally, the prediction of ultimate status is less certain. In two cases known to be children from foreign speaking homes, the scores appear to have improved as the children attained better command of English.

TABLE III.—DISTRIBUTION OF GAINS AND LOSSES IN PERCENTILE RANK, ACE AND BINET TEST RESULTS

Gain or loss	Frequency	Boys	Girls
50	1	1	
45	2	1	1
40	5	2	3
35	3	2	1
30	3	2	1
25	2	1	1
20	2	1	1
15	4	0	4
10	5	2	3
5	4	1	3
0	11	7	4
− 5	9	7	2
−10	6	1	5
−15	4	1	3
−20	4	1	3
−25	0		
−30	2	1	1
−35	0		
−40	1		1
−45	1	1	
−50	1		1
−55			
Total.....	70	32	38

Median change in percentile rank ACE compared with Binet = 3.2 points
Q_1 = −9.8
Q_3 = 21.2
Median change boys = 2.9 points
Median change girls = 3.75 points

In general, the results show for the population as a whole more gain than loss in centile status. The tendency for improvement in centile status to be larger than decrease in centile status may be a chance finding or it may be attributable to differences in the extreme ends of the two populations used for converting the test scores into percentile ranks. Subsequent studies may prove that this tendency is attributable to the nature of the school population itself.

TABLE IV.—TABLES USED FOR CONVERTING BINET IQ'S AND ACE SCORES INTO PERCENTILE RANK

Binet IQ	Lincoln school percentile	Binet IQ	Lincoln school percentile	ACE scores	Percentile	ACE scores	Percentile
170	99						
169	99	119	46	355	100	145	12
168	99	118	44	350	100	140	10
167	99	117	40	345	100	135	8
166	99	116	36	340	99	130	7
165	99	115	32	335	99	125	7
164	99	114	30	330	99	120	6
163	99	113	26	325	98	115	5
162	99	112	24	320	98	110	3
161	99	111	19	315	97	105	2
160	99	110	16	310	96	100	1
159	99	109	15	305	96	95	1
158	99	108	12	300	95	90	1
157	99	107	10	295	94		
156	99	106	8	290	94		
155	99	105	7	285	93		
154	98	104	5	280	92		
153	98	103	4	275	90		
152	98	102	3	270	88		
151	98	101	3	265	87		
150	97	100	2	260	85		
149	97	99	2	255	83		
148	96	98	2	250	81		
147	96	97	2	245	79		
146	95	96	2	240	77		
145	95	95	1	235	74		
144	94	94	1	230	70		
143	94	93	1	225	67		
142	93	92	1	220	62		
141	92	91	1	215	57		
140	92	90	1	210	52		
139	91	...	..	205	48		
138	90	...	..	200	44		
137	88	...	..	195	41		
136	87	...	..	190	37		
135	86	...	..	185	33		
134	85	...	..	180	30		
133	82	...	..	175	28		
132	80	...	..	170	25		
131	78	...	..	165	23		
130	76	...	..	160	20		
129	75	...	..	155	17		
128	71	...	..	150	14		
127	70						
126	67						
125	64						
124	61						
123	58						
122	55						
121	51						
120	48						

BIBLIOGRAPHY

1. Cattell, Psyche: "Do the Stanford-Binet IQ's of superior boys and girls tend to decrease or increase with age?" *Journal of Educational Research*, Vol. XXVI, May, 1933, pp. 668–673.
2. Hollingworth, Leta S., and Kawitz, R. M.: "The centile status of gifted children at maturity." *Journal of Genetic Psychology*, Vol. XLV, October, 1934, pp. 106–120.
3. Lincoln, E. A.: "A Study of changes in intelligence quotients of superior children." *Journal of Educational Research*, Vol. XIX, December, 1935, pp. 272–275.
4. Lorge, Irving, and Hollingworth, Leta S.: "Adult status of highly intelligent children." *Pedagogical Seminary and Journal of Genetic Psychology*, Vol. XLIX, September, 1936, pp. 215–226.
5. Terman, L. M., and Others: "The Promise of Youth." *Genetic Studies of Genius*, Vol. III. Stanford University, California: Stanford University Press, 1931.
6. Wellman, Beth L.: "Mental growth from pre-school to college." *Journal of Experimental Education*, Vol. VI, December, 1937, pp. 127–138.

SOME OBSERVATIONS CONCERNING THE RELATIONSHIP OF TIME SPENT AT STUDY TO SCHOLARSHIP AND OTHER FACTORS

D. G. RYANS

William Woods College

May,[3] Jones and Ruch,[2] Crawford,[1] and Williamson[6] have reported results obtained when the number of hours spent at study by university students are considered in relation to estimates of scholarship and mental test results. Findings of the several investigations have been reviewed and compared by Williamson.[6] It is the aim of the present discussion to take further cognizance of study time as a variable of psychological as well as educational interest and significance. Data with respect to the study activity of a group of junior-college sophomores under uncontrolled (study for college courses during one week) and controlled (study at a limited assignment, under supervision) conditions will be introduced.

Forty subjects, members of the writer's classes in General Psychology at the Eveleth Junior College, voluntarily agreed to participate in the study. They were unaware of the ultimate purpose of the experiments. No intentional selection was exercised regarding intelligence, school marks, or estimated conscientiousness and industry. The sexes were equally represented. Ages of the subjects ranged from eighteen to twenty-five, the mean falling just under twenty years. Scores on the Revised Army Alpha Examination extended from 80 to 203; on the Otis Self-Administering Test, Higher Examination, from 28 to 69. Means and standard deviations for the two tests were 148.0 ± 21.15 and 47.65 ± 8.16, respectively. Honor point ratios (based upon the following allotment of points: $A = 3$; $B = 2$; $C = 1$; $D = 0$; $F = -1$) averaged 1.32, SD, 0.6.

In all, six sets of measurements were employed in the investigation: Study ratios (the recorded number of minutes spent at study during one week divided by the number of credit hours comprising an individual course load); controlled study times (the recorded number of minutes spent at study on a limited assignment); honor point ratios (the number of honor points earned during the semester divided by the number of credit hours comprising an individual course load); intelligence indices (derived by equating Army Alpha Examination and Otis

Test scores and determining the typical rating for each individual); persistence scores (obtained through the administration of a persistence test (4) previously developed); and work habits inventory scores (obtained through the administration of a schedule prepared by the writer).

Study ratios were calculated from study logs. Each subject was required to keep an inclusive record of the time devoted to studying for college courses during one full week. Mimeographed record blanks were distributed in order to facilitate the task. Upon each blank were printed the following specific directions:

> Keep a record of all the time that you spend studying. Be accurate! *This has no bearing on your grades.* Be honest and do your best to make this record correct. Keep a record for each day for one week. (Dates specified.) In doing this, indicate (1) time spent in study, (2) place where studying was done—home, study hall, library, etc.—(3) subject studied, and (4) whether you studied alone or with someone else.*

A score consisted of the total number of minutes studied during the week, divided by the number of academic credit hours which made up the student's program. The average score was 48.65, with a standard deviation of 18.25.

For the controlled study time experiment, five pages of mimeographed expository materials (available to the students in no other form) were distributed in class. Instructions specified that the contents were to be studied for substance, and that an examination covering them would be administered at the following meeting of the groups. Each individual was allowed as much time as he pleased, or thought necessary, for study. Whenever a subject believed he had attained familiarity with the material, he was excused for the period. It was required that no notes be taken and that all materials be deposited with the instructor at the conclusion of study. As a set of papers was returned, the study time of the bearer was recorded. A score consisted simply of the number of minutes, or fractions thereof, devoted to preparation. The average time spent was 44.75 minutes, while the standard deviation was 3.52.

All measurements obtained in the investigation were translated into standard deviation scores with means of 50 and standard deviations of 10, to make possible the ready comparison of results.

---

* The last three items were included in an effort to secure an accurate report. They were not scored.

As judged from the statistical treatment of available data time spent at study, while, in general, unrelated to intelligence, is an important factor in determining the success or failure of a student in his academic endeavors. Table I shows the product-moment coefficients of correlation revealed when the variables—weekly study ratio, intelligence, persistence, honor-point ratio, and professed work habits —are compared. Study-time appears to be positively related to scholarship, as scholarship is portrayed by marks in junior-college courses.

TABLE I.—INTERCORRELATION COEFFICIENTS INDICATING THE EXTENT TO WHICH VARIATIONS IN WEEKLY STUDY TIME, INTELLIGENCE, PERSISTENCE, SCHOLARSHIP, AND WORK HABIT INVENTORY SCORES WERE FOUND TO BE COINCIDENT
(N = 40)

	Intelligence	Persistence	Honor-point ratio	Work habit inventory
Weekly study ratio.....	−0.06	0.56	0.37	0.47
Intelligence.....	.....	−0.13	0.48	0.01
Persistence.....	.....	.....	0.48	0.46
Honor-point ratio.....	.....	.....	....	0.29

In the above table, it may be observed that a questionnaire regarding the possession of effective work habits seemed to be significantly symptomatic of the amount of time spent at study. Each of these measures, (weekly study ratio and work habits inventory), in turn, was found to correlate positively with the results of a persistence test administered. In light of these inter-relationships it appears probable that study time and the questionnaire may be reflecting, directly and indirectly, a more basic behavior tendency known as perseverance. In Table II, this relationship between study ratio scores and persistence scores is further substantiated through the indication of the significance of the difference between high and low persistence groups observed with respect to weekly study times.

Persistence has been considered in an earlier discussion[5] as a factor of importance in the determination of academic marks. Positive correlations between intelligence and school success, and between persistence and school success, along with almost complete lack of

relationship between persistence and intelligence were observed to contribute to a multiple coefficient of correlation between scholarship, on the one hand, and intelligence and persistence combined, on the other, of approximately 0.80.

TABLE II.—A COMPARISON OF STUDY RATIO SCORES* WITH RESPECT TO PERSISTENCE TEST† RESULTS
(N = 40)

	Average study ratio score	SD	Diff./$SD_{(Diff.)}$	Per cent scores above SD score of fifty	Per cent scores below SD score of fifty
Highest quarter of persistence distribution....	58.7	6.8	...	70	30
Lowest quarter of persistence distribution....	45.2	9.5	3.7	10	90

* Study ratios (the recorded number of minutes spent at study during one week divided by the number of credit hours comprising an individual course load) were translated into standard deviation scores with a mean of fifty and standard deviation of ten.

† Ryans, D. G.: "An experimental attempt to analyze persistent behavior. II. A persistence test." *Jour. Gen. Psychol.*, Vol. XIX, 1938, pp. 355–371.

Table III, partially adapted from a similar presentation by Williamson,[6] summarizes the correlational results of several studies having to do with study time and its influences on college marks. Our data obtained from weekly study records bear closest resemblance to the larger sample studied at Syracuse University by May.[3] Scholarship and intelligence, and scholarship and study-time are positively correlated, while intelligence and study-time show no statistical relationship.

It is interesting to note that time spent at study in a controlled situation with limited materials yields correlation coefficients with the other variables quite similar, though less significant, to those obtained when weekly records of study were considered. The multiple coefficients, predicting academic marks from the combined measures of intelligence and study-time, are 0.63 and 0.52, respectively, when weekly and limited study times are considered.

SUMMARY

Times spent at study for college classes during one week and times spent at study on a limited assignment were observed in relation to scholastic success as indicated by course marks. Both variables were found to correlate positively with scholarship, and both were lacking in evident relationship to intelligence. It is concluded that the present study verifies the work of at least one previous investigator[3] in noting study time to be a significant factor in the earning of grades. (Other attempts to clarify this problem[1,2,6] sometimes have resulted in contrary findings.) An intimate relationship between persistence scores and study times suggests that the degree of perseverance possessed by an individual may determine, to an extent, how much he will study and, in turn, partially, his scholarship status.

TABLE III.—COEFFICIENTS OF CORRELATION INDICATING THE EXTENT TO WHICH (1) COLLEGE MARKS, (2) INTELLIGENCE INDICES, AND (3) TIMES SPENT AT STUDY HAVE BEEN OBSERVED TO VARY CONCOMITANTLY

	Time spent at study during one week					Time spent at study on a limited assignment, junior college sophomores (N = 40)
	University freshmen*				Junior college sophomores (N = 40)	
	Syracuse (N = 450)	Yale (N = 221)	Minnesota (N = 105)	Iowa (N = 130)		
r_{12}	0.60	0.28	0.65	0.69	0.48	0.48
r_{13}	0.32	0.00	−0.06	−0.28	0.37	0.21
r_{23}	−0.35	−0.15	−0.20	−0.41	−0.06	0.00
$R_{1.23}$	0.82	0.28	0.66	0.69	0.63	0.52

* Adapted from Williamson, E. G.: "The relationship of number of hours of study to scholarship." *Jour. Educ. Psychol.*, Vol. XXVI, 1935, pp. 682–688.

BIBLIOGRAPHY

1. Crawford, A. B.: *Incentives to Study.* New Haven, Conn., Yale University Press, 1929.
2. Jones, L. and Ruch, G. M.: "Achievement as affected by amount of time spent in study." *Twenty-seventh Yearbook*, National Society for the Study of Education. Part V, 1928, pp. 131–134.

3. May, M. A.: "Predicting academic success." *Jour. Educ. Psychol.*, Vol. XIV, 1923, pp. 429–440.
4. Ryans, D. G.: "An experimental attempt to analyze persistent behavior. II. A persistence test." *Jour. Gen. Psychol.*, Vol. XIX, 1938, pp. 355–371.
5. Ryans, D. G.: "A study of the observed relationship between persistence test results, intelligence indices, and academic success." *Jour. Educ. Psychol.*, Vol. XXIX, 1938, pp. 573–580.
6. Williamson, E. G.: "The relationship of number of hours of study to scholarship." *Jour. Educ. Psychol.*, Vol. XXVI, 1935, pp. 682–685.

CERTAIN FACTORS RELATED TO LIBERAL AND CONSERVATIVE ATTITUDES OF COLLEGE STUDENTS: SEX, CLASSIFICATION, FRATERNITY MEMBERSHIP, MAJOR SUBJECT

PAUL J. FAY AND WARREN C. MIDDLETON

DePauw University

Psychologists and educators are to an increasing extent recognizing the significance of attitudes as factors in personality. Many problems, however, must yet be solved before the conscious fostering of desirable attitudes can become anything more than a pious wish. The development and improvement within recent years of instruments for the measurement of attitudes have made possible an effective attack upon some of these problems. Numerous studies have been made of the existing attitudes of certain groups. Some information has been acquired regarding the stability and permanence of attitudes. But comparatively little is known about their genesis and the factors which condition them. The specific purpose of this study is to discover to what extent certain factors with which colleges are especially concerned are related to some of the liberal and conservative attitudes of college students.[1] The study is part of a larger investigation which we are making in an effort to discover some factors that may be significant in the genesis and conditioning of attitudes.

Five of the Thurstone, *et. al.*, Scales for *The Measurement of Social Attitudes* were used: Form *A* of each of the scales measuring attitudes toward Communism, Patriotism, the United States Constitution, Law, and Censorship. We selected these particular attitudes for study, because they seem to constitute some of the elements of political liberalism and conservatism, particularly in relation to national

[1] We are reporting the extent to which certain other factors are related to liberal and conservative attitudes as follows:

(1) "Certain factors related to liberal and conservative attitudes of college students: Parental membership in certain organizations." *Journal of Social Psychology*, in press.

(2) "Certain factors related to liberal and conservative attitudes of college students: I. Father's occupation; size of home town." *Journal of Social Psychology*, in press.

(3) "Certain factors related to liberal and conservative attitudes of college students: II. Father's political preference; presidential candidates favored in the 1932 and 1936 elections." *Journal of Social Psychology*, in press.

politics. The five scales were clipped together with a personal history questionnaire. To encourage frankness of response the subjects were instructed not to sign their names to the blanks.

The subjects were five hundred seventy-five students of DePauw University, nearly one-half of the total enrollment. Two hundred seventy-eight were men and two hundred ninety-seven were women. All classes were represented, as follows: Freshmen, one hundred eighty-five; sophomores, one hundred sixty-eight; juniors, one hundred eight; seniors, one hundred fourteen. Four hundred sixty-one were members, or pledges, of social fraternities; one hundred fourteen were unorganized. Although this is a larger number of subjects than is used in most attitude studies, it is a random sampling neither of the total population nor of college students in general. The great majority of the subjects come from the States of Indiana, Illinois and Ohio. They may be a fairly typical sample of college students from upper middle-class homes in these States. They form an unusually homogeneous group. Nearly all are members of one or another of the evangelical denominations; very few Catholics, Jews, or non-church members are included. The fathers of only fifteen of the subjects are manual laborers, and of only twenty are farmers. The distribution of fathers among other occupational groups is as follows: Proprietoral, one hundred fifteen; professional, one hundred sixty-three; managerial, one hundred sixty-six; commercial, forty-eight; clerical, twenty-eight. The number of subjects from communities of various sizes is as follows: In the country, twenty-six; up to two thousand five hundred, fifty; two thousand five hundred one to ten thousand, eighty-six; ten thousand one to twenty-five thousand, seventy-five; twenty-five thousand one to one hundred thousand, one hundred twenty-six; one hundred thousand one to five hundred thousand, one hundred twenty-four; over five hundred thousand, eighty-three.

Additional evidence concerning the family background of the subjects is found in the distribution of political preferences of their fathers. These preferences are as follows: Republican, four hundred twenty-eight; Democratic, one hundred eighteen; Socialist, six; Union, one; Communist, none. The presidential candidates favored by the subjects themselves in the 1936 election were: Landon, three hundred eighty-seven; Roosevelt, one hundred sixty-seven; Thomas, eleven; Lemke, none; Browder, none. The two newspapers most frequently listed by the subjects as regularly taken in their homes were the *Indianapolis Star* and the *Chicago Tribune.*

The relationship between the sex of the subjects and their scores on the various attitude scales is summarized in Table I. The data indicate that the men are more liberal than the women students in all the attitudes measured except the attitude toward patriotism. For the purpose of this study it is assumed that a high score on the communism scale and a low score on the other scales represents liberalism; the converse represents conservatism. We shall not attempt to defend this use of the terms "liberalism" and "conservatism" except to point out that apparently it agrees with popular usage. The median for men students on the communism scale is 5.43. One can hardly assert that 5.43 on an eleven-point scale represents a radical attitude.

TABLE I.—MEDIAN, MEAN, AND SD FOR MALE AND FEMALE COLLEGE STUDENTS, DIFFERENCES BETWEEN THE MEDIAN SCORES OF THE SEXES, AND RELIABILITY OF THE DIFFERENCES ON FIVE DIFFERENT SOCIAL ATTITUDE SCALES

Attitude scale	Male				Female				Difference[1] between medians		$\sigma_{diff.}$	$\frac{D}{\sigma_{diff.}}$
	N	Median	Mean	SD	*N*	Median	Mean	SD	Male	Female		
Communism...	278	5.43	4.87	1.26	297	4.33	4.54	1.14	1.10	...	.13	8.46
Patriotism.....	278	6.33	6.34	.99	297	6.17	6.23	.99	.16	...	.10	1.60
Constitution...	278	5.82	5.84	.89	297	6.02	5.97	.92	....	.20	.09	2.22
Law.....	278	5.94	6.26	.65	297	6.16	6.18	.63	....	.22	.06	3.67
Censorship.....	278	5.51	5.17	1.43	297	5.76	5.39	1.35	....	.25	.15	1.67

[1] The difference between the median scores of the two sexes is printed in the column of the sex which has the higher median.

At the most it indicates a less intolerant attitude toward communism than does the women's median. The difference of 1.10 between the medians of the two sexes is one-tenth of the scale's range. Since the difference is 8.46 times as large as its standard error, it is decidedly reliable. The medians for the men represent less favorable attitudes toward the Constitution, law, and censorship than do the women's medians. None of these sex differences are as large as the difference in attitude toward communism; nor are they as reliable. There are, however, at least ninety-five chances in one hundred that the men are more liberal than the women in each of these three attitudes.

Greater differences are found between students from different classes, as shown in Table II. The frequently expressed criticism of

TABLE II.—MEDIAN, MEAN, AND SD FOR COLLEGE STUDENTS OF EACH CLASS, DIFFERENCES BETWEEN THE MEDIAN SCORES OF FRESHMEN AND SENIORS, AND RELIABILITY OF THE DIFFERENCES ON FIVE DIFFERENT SOCIAL ATTITUDE SCALES

Class	*N*	Communism			Patriotism			Constitution			Law			Censorship		
		Median	Mean	SD	Median	Mean	SD	Median	Mean	SD	Median	Mean	SD	Median	Mean	SD
Freshmen	185	4.37	4.48	1.19	6.47	6.47	1.11	6.24	6.14	.88	6.19	6.30	.66	6.03	5.45	1.41
Sophomores	168	4.34	4.54	1.18	6.41	6.38	.96	6.08	6.00	.86	5.95	6.12	.74	5.62	5.16	1.38
Juniors	108	4.80	4.82	.98	6.23	6.23	.83	5.63	5.71	.64	5.91	5.98	.62	5.62	5.28	1.38
Seniors	114	5.27	5.13	1.35	5.84	5.93	.89	5.44	5.53	.92	5.87	5.98	.64	4.79	4.94	1.40
Difference[1] between:																
Freshmen	....	....	....	....	.63	....	....	.80	....	....	.32	....	....	1.24		
Seniors	....	.90														
$\sigma_{diff.}$	....	.19	....	....	.15	....	....	.13	....	....	.10	....	....	.21		
$D/\sigma_{diff.}$	....	4.74	....	....	4.20	....	....	6.15	....	....	3.20	....	....	5.90		

[1] The difference between the median scores of freshmen and seniors is printed in the same row as the name of the group having the higher median. Seniors have the higher median on the communism scale, freshmen on the other scales.

colleges, that their students become increasingly less patriotic, seems to be substantiated. The trend, however, can scarcely be called menacing, since there is a drop in class median during the four years of only .63. The typical senior has reached the indifference point in his attitude toward patriotism; he has not become antagonistic to it. Although seniors are .90 more tolerant than freshmen toward whatever they conceive communism to be, they are still on the antagonistic side of indifference. The decline of .32 in favorable attitude toward law is the least for any of the five attitudes. The greatest drop in class medians is in attitude toward censorship: It declines from the indifference of freshmen to the 4.79 of seniors. The median of the seniors, however, still represents considerable tolerance toward censorship for individuals who have been subjected for nearly four years to a liberal education. There are one hundred chances in one hundred that seniors are more liberal than freshmen in each of the five attitudes measured. It is likewise significant that, when both measures of central tendency are considered, there is a consistent increase each year in liberalism in all five attitudes.

The data do not reveal whether or not the increased liberalism is due to any factors directly related to the college itself. Some other factors that may be contributory are: Greater maturity; greater freedom from parental influence; increased experience; and, as a result of legal maturity, greater political interest.

The differences between the attitudes of members, or pledges, of fraternities, and the attitudes of unorganized students are shown in Table III. The median of non-fraternity members is 1.00 more favorable toward communism than is the median of fraternity members. As far as the other attitudes are concerned, there is very little difference between organized and unorganized students. The belief, sometimes expressed, that fraternity members are much more conservative is not borne out by this study. In a state or urban university a greater difference between the two groups of students might be revealed.

The medians, means, and SD's on each of the five attitude scales for students majoring in various college subjects are shown in Table IV. Students majoring in physics, chemistry, geology, botany, and zoology are combined in the biological and physical science group. Students majoring in Latin, French, Spanish, and German are combined in the foreign language group. Those majoring in philosophy and in

English Bible are combined in one group, and those majoring in art and music in another group.

Philosophy majors are the most favorable of any toward communism; the median of the group is 5.62. Students majoring in political science with a median of 5.50 closely approach the philosophy group

TABLE III.—MEDIAN, MEAN, AND SD FOR FRATERNITY AND NON-FRATERNITY COLLEGE STUDENTS, DIFFERENCES BETWEEN THE MEDIAN SCORES OF THE GROUPS, AND RELIABILITY OF THE DIFFERENCES ON FIVE DIFFERENT SOCIAL ATTITUDE SCALES

Attitude scale	Fraternity				Non-fraternity				Difference[1] between medians		$\sigma_{diff.}$	$\frac{D}{\sigma_{diff.}}$
	N	Median	Mean	SD	*N*	Median	Mean	SD	Fraternity	Non-fraternity		
Communism...	461	4.00	4.61	1.20	114	5.00	5.12	1.19	...	1.00	.16	6.25
Patriotism.....	461	6.29	6.29	.99	114	6.11	6.21	.91	.18	....	.12	1.50
Constitution...	461	5.92	5.93	.98	114	5.89	5.81	.95	.03	....	.12	.25
Law.....	461	5.98	6.13	.67	114	5.96	6.11	.70	.02	....	.09	.22
Censorship.....	461	5.65	5.22	1.42	114	5.68	5.52	1.23	...	.03	.17	.18

[1] The difference between the median scores of the fraternity and non-fraternity students is printed in the column of the group which has the higher median.

in tolerance toward communism. The groups most opposed to communism are majors in art and music (3.89), home economics (4.25), and psychology (4.31). There may be some significance in the fact that majors in the two subject-matter fields in which there is the greatest likelihood of acquiring some understanding of the philosophy of communism reveal the greatest tolerance for it. The gap of .62 between the median of political science majors and that of science majors, the third most liberal group, is by far the largest break in the series of medians on this scale.

Students majoring in philosophy are likewise most liberal in their attitude toward patriotism. Foreign language majors are the second lowest in favorable attitude toward patriotism; they are more liberal in this attitude than in any other of the five. The better appreciation of other countries that is fostered by foreign language study may make an individual more critical of the nationalistic shibboleths of his own country. The range of the medians on this scale is comparatively

TABLE IV.—MEDIAN, MEAN, AND SD OF SCORES ON FIVE DIFFERENT SOCIAL ATTITUDE SCALES FOR COLLEGE STUDENTS MAJORING IN VARIOUS SUBJECTS

Major subject	*N*	Communism			Patriotism			Constitution			Law			Censorship		
		Median	Mean	SD	Median	Mean	SD	Median	Mean	SD	Median	Mean	SD	Median	Mean	SD
Philosophy and Bible	10	5.62	5.55	.68	5.80	6.05	.52	5.66	5.60	.76	5.75	5.70	.49	5.75	4.69	1.89
Biological and physical sciences	106	4.88	4.80	1.10	6.11	6.18	.97	5.73	5.77	.86	5.97	5.78	.55	4.81	4.98	1.42
Sociology	43	4.64	4.83	1.21	5.97	6.17	1.15	5.33	5.46	.85	5.90	5.91	.73	6.00	5.50	1.38
Foreign languages	39	4.85	4.76	1.15	5.97	6.10	.91	5.88	5.90	.76	5.96	6.12	.63	5.00	5.07	.54
Political science	34	5.50	5.37	1.22	6.50	6.60	.96	6.00	5.92	.79	5.95	6.15	.68	4.30	4.72	1.35
English	97	4.86	4.79	1.17	6.10	6.18	.89	5.87	5.86	.95	6.03	6.13	1.36	5.79	5.38	1.24
Economics	94	4.65	4.60	1.25	6.56	6.57	.95	5.80	5.83	.96	5.97	6.12	.72	4.90	5.06	1.45
History	38	4.81	4.78	1.06	6.25	6.14	1.23	6.16	6.01	.99	5.99	6.34	.92	5.68	5.29	1.50
Psychology	25	4.31	4.59	1.22	5.99	6.23	.89	6.10	6.04	.90	5.94	6.01	.52	5.80	5.54	1.14
Art and music	50	3.89	4.27	1.14	6.29	6.36	.84	6.37	6.27	.83	6.05	6.16	.79	6.33	5.91	1.35
Home economics	30	4.25	4.50	.97	6.62	6.69	.96	6.53	6.20	1.05	6.12	6.35	.53	5.83	5.50	1.30

small: Only .82 separates philosophy majors (the least patriotic group) from home economics majors (the most patriotic group). In contrast, on the communism scale the philosophy group is 1.73 higher than the art and music group.

The sociology group reveals the most liberal attitude toward the Constitution. Its median of 5.33 is .33 lower than that of the second-place philosophy majors, and 1.20 lower than the median of the most conservative group, home economics majors. It is somewhat surprising to find history majors with a median of 6.16 closely following the home economics and art and music groups in their conservative attitude toward the Constitution. Students majoring in political science are relatively conservative in their attitudes toward the Constitution and patriotism.

On the scale measuring attitude toward law the medians of the groups range from .12 above the indifference point (6.00) to .25 below it. Students majoring in home economics reveal the most favorable attitude and those majoring in philosophy the least favorable.

The range of group medians on attitude toward censorship is 2.03; this is the largest of any in the study. The political science group with a median of 4.30 is most opposed to censorship. Its median is .51 lower than that of the second most liberal group, science majors. Students majoring in political science are probably somewhat more familiar than other students with the Bill of Rights. Art and music majors with a median of 6.33 reveal the most favorable attitude toward censorship. For some unaccountable reason sociology majors, who are very liberal on the other four scales, are with respect to censorship the second most conservative group.

The variability of the scores for the various subject-matter groups differs considerably from scale to scale. The greatest heterogeneity in attitude occurs toward censorship: The mean of the SD's of the eleven groups is 1.32. The second greatest heterogeneity is found in the attitude toward communism with a mean SD of 1.11. The subjects are most homogeneous in their attitude toward law: The mean SD on this scale is .72. The mean SD's for the patriotism and Constitution scales are, respectively, .93 and .88. The various groups likewise differ in the variability of the scores within each group. Foreign language majors with a mean SD of .80 for the five scales are the most homogeneous. Students majoring in history with a mean SD of 1.16 are the most heterogeneous.

TABLE V.—DIFFERENCES BETWEEN THE MEDIAN SCORES ON FIVE DIFFERENT SOCIAL ATTITUDE SCALES AND RELIABILITY OF THE DIFFERENCES FOR COLLEGE STUDENTS MAJORING IN VARIOUS SUBJECTS

Major subjects compared	Communism			Patriotism			Constitution			Law			Censorship		
	Difference[1]	$\sigma_{diff.}$	$\frac{D}{\sigma_{diff.}}$	Difference[1]	$\sigma_{diff.}$	$\frac{D}{\sigma_{diff.}}$	Difference[1]	$\sigma_{diff.}$	$\frac{D}{\sigma_{diff.}}$	Difference[1]	$\sigma_{diff.}$	$\frac{D}{\sigma_{diff.}}$	Difference[1]	$\sigma_{diff.}$	$\frac{D}{\sigma_{diff.}}$
Philosophy	1.73	.34	5.09	.49	.25	1.96	.71	.34	2.09	.30	.24	1.25	.58	.79	.73
Art, music	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....
Philosophy	1.37	.35	3.91	.82	.30	2.72	.87	.39	2.23	.37	.23	1.61	.08	.80	.10
Home economics	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....
Philosophy	.99	.24	4.13	.18	.19	.95	.64	.18	3.56	.08	.15	.53	1.52	.30	5.07
Sciences	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....
Art, music	.63	.26	2.42	.51	.25	2.04	.80	.26	3.08	.15	.14	1.07	1.02	.34	3.00
Sciences	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....
Home economics	.75	.30	2.50	.32	.26	1.23	1.04	.21	4.95	.15	.20	.75	.33	.35	.94
Sociology	....	....	....	....	....	....	....	....	....	....	....	....	.17	....	....
Art, music	.39	.33	1.18	.65	.31	2.11	1.20	.29	4.14	.22	.19	1.16	....	.40	.43
Sociology	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....
Home economics	1.61	.33	4.88	.21	.25	.84	.37	.23	1.61	.10	.20	.50	2.03	.38	5.34
Political science	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....
Art, music	1.25	.34	3.68	.12	.30	.40	.53	.29	1.83	.17	.19	.89	1.53	.41	3.72
Political science	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....
Home economics	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....

[1] The difference between the median scores of the two groups of students is printed in the same row as the name of the major subject whose students have the higher median. In comparing, for example, students majoring in philosophy and Bible with those majoring in art and music, philosophy majors have a higher median on the communism scale; art and music majors have higher medians on the other four scales.

Differences between the median scores and reliabilities of the differences for some of the groups on each of the five scales are shown in Table V. Each of the three most liberal groups and the political science group are compared with the art and music group and the home economics group; the last two groups are the most conservative. Most of the differences shown between group medians on the communism scale are more than three times as large as their standard errors. Four of the differences on the censorship scale and four on the Constitution scale are equally reliable.

In Table VI the groups of students majoring in the various subjects are ranked for liberalism in the five attitudes measured. The total ranks are probably of some significance for the groups at the extreme top, and for those at the extreme bottom, of the table. Since students majoring in philosophy and Bible and in the sciences are consistently liberal in all the five attitudes, it is probably justifiable to conclude that, as far as these attitudes are concerned, these two groups are the most liberal in college. For similar reasons it seems probable that majors in home economics, art, and music are the most conservative. Sociology majors are liberal in all attitudes except that toward censorship. Students majoring in political science, on the other hand, are liberal in their attitudes toward censorship and communism, but are relatively conservative in the other attitudes. The fact that most of the groups are fairly consistent in their ranks on the different scales offers some justification for the use of some such general terms as "liberal" and "conservative."

The data supply no evidence concerning the cause or causes of the differences in attitude between the various groups of students. It does not necessarily follow from the results that instruction in philosophy produces more liberal attitudes in students than does instruction in home economics. The students included in any one major group have taken varying numbers of courses in that field. Some freshmen and sophomores may have declared, or selected, a major without having taken any courses in that field. Some students may, at the time the data were collected, actually have had more hours of credit in some department other than the one in which their major was declared. It is quite possible that different instructors in a given department foster in students different degrees of liberalism. Various selective factors may operate in such a way as to account for some of the differences between subject groups. It is possible, for example,

TABLE VI.—RANK IN LIBERALISM AS DERIVED FROM MEDIANS AND MEANS ON EACH OF FIVE DIFFERENT SOCIAL ATTITUDE SCALES FOR COLLEGE STUDENTS MAJORING IN VARIOUS SUBJECTS

Major subject	Rank in liberalism[1]													
	Communism		Patriotism		Constitution		Law		Censorship		Total			
	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Combined[2]
Philosophy and Bible	1	1	1	1	2	2	1	1	6	1	1	1	1
Biological and physical sciences	3	4	6	5.5	3	3	6.5	2	2	3	2	2	2
Sociology	8	3	2.5	4	1	1	2	3	10	8.5	5	3	3
Foreign languages	5	7	2.5	2	6	6	5	5.5	4	5	3	4	4
Political science	2	2	9	10	7	7	4	8	1	2	4	5	5
English	4	5	5	5.5	5	5	9	7	7	7	6	6	6
Economics	7	8	10	9	4	4	6.5	5.5	3	4	7	7	7
History	6	6	7	3	9	8	8	10	5	6	9	8	8
Psychology	9	9	4	7	8	9	3	4	8	10	8	9	9
Art and music	11	11	8	8	10	11	10	9	11	11	10	10	10
Home economics	10	10	11	11	11	10	11	11	9	8.5	11	11	11

[1] The higher a score is on the communism scale, the more liberal it is assumed to be; the lower a score is on the patriotism, Constitution, law, and censorship scales, the more liberal it is assumed to be.

[2] Rank derived from the sum of the ranks in both median and mean columns.

that students who are already fairly liberal are more likely to major in an abstract and intellectualistic subject, such as philosophy, and that conservative students are more likely to major in a concrete and applied subject, such as home economics. In our opinion the factor of selection partially explains the general conservative attitude of students majoring in psychology. The majority of this group are students who are particularly interested in applications of psychology, such as personnel work, clinical work, and advertising. Certainly, a study of pure psychology is unlikely to foster in students acceptance of the popular shibboleths and slogans that constitute the statements in the conservative extremes of the five scales used.

In any case the problem seems worthy of better controlled study in the future. Similar studies should be made in institutions with more heterogeneous student bodies. If information concerning the number of hours that students had taken in each department were recorded, the partial correlation technique might profitably be used. By administering one form of each of the scales before students had taken any courses in a given field, and another form after they had taken one or more courses, selective factors and specific influences of the courses might be determined. Some such information is needed before educational institutions, or particular departments, can substantiate some of their current claims concerning their influence on the attitudes of students.

The conclusions drawn from the present study—conclusions that are highly tentative and that are applicable only to the subjects included in the study—may be summarized as follows:

(1) College men tend to be more liberal than college women in their attitude toward communism, the United States Constitution, law, and censorship; the sex difference in liberalism is most pronounced in attitude toward communism.

(2) There is a small, but consistent, increase each year from freshmen to seniors in liberalism in all five attitudes; the differences between the attitudes of freshmen and seniors are statistically reliable.

(3) Unorganized students are considerably more tolerant toward communism than are fraternity members or pledges; in the other four attitudes there exist only slight differences between members and non-members of fraternities.

(4) In the five attitudes studied students who major in philosophy, Bible, and the sciences are the most liberal of any groups of majors.

(5) In the five attitudes studied students who major in home economics, art, and music are the most conservative.

(6) In general, groups of students who major in any one field are fairly consistent in their liberalism or conservatism in all five attitudes studied.

(7) College students vary most in their attitude toward censorship and vary least in their attitude toward law.

CONVERTING PERCENTAGE DIFFERENCES INTO TETRACHORIC CORRELATION COEFFICIENTS

SAMUEL P. HAYES, JR.

Chicago, Illinois

It is often desirable to get a quick indication of the relationship between two variables. Where the greatest reliability is not necessary, and especially where one—or both—of the variables has been tabulated into two classes only, the method of tetrachoric correlation is particularly well adapted to this purpose. Certain tables[1] and diagrams[2] are already available for the rapid calculation of the tetrachoric coefficients. The present table is intended for the same purpose, but the data needed to enter it are of a slightly different nature from the other available short-cuts; the table is more convenient for some persons to use and, in addition, it contains a column from which the probable errors of the coefficients can be readily determined.

Because of the convenience of speaking of percentage differences between groups when any of a wide variety of variables is being discussed, and because of the ease of calculating them, the table has been calculated and arranged with reference primarily to such percentage differences. However, it can also be used to obtain probable errors for tetrachoric coefficients obtained in any other way. In such a case, the percentage difference is disregarded.

Of course, in order to use this table, such data as are not already grouped into two classes must be so grouped. The criteria in terms of which the variables are grouped should be chosen so that very unequal divisons do not result. Otherwise, the coefficients obtained will be relatively unreliable and will demand more difficult interpolation. In addition, it must be remembered that the tetrachoric method is only valid where there are reasonable grounds for assuming that both variables are continuous, and that both show distributions approaching normality. Markedly skewed curves will not change the coefficients very seriously (for such unreliable measures) if the dichotomies are not extreme, but if the dichotomies for skewed curves are extreme, or if the variables are not continuous, the resulting coefficients will be practically meaningless.

---

[1] Pearson, K., *Editor: Tables for Statisticians and Biometricians*, Parts 1 and 2. Cambridge University Press.

[2] Cheshire, L., Saffir, M., and Thurstone, L. L.: *Computing diagrams for the tetrachoric correlation coefficient.* University of Chicago Bookstore, 1933.

TABLE FOR CONVERSION OF PERCENTAGE DIFFERENCE BETWEEN GROUPS INTO COEFFICIENTS OF TETRACHORIC CORRELATION

Percentage of whole group above criterion	Relative sizes of compared subgroups	Difference of 5 per cent		Difference of 10 per cent		Difference of 15 per cent		Difference of 20 per cent	
		r_t	PE$(\sqrt{N})$	r_t	PE$(\sqrt{N})$	r_t	PE$(\sqrt{N})$	r_t	PE$(\sqrt{N})$
50	50–50	.07	1.05	.15	1.04	.23	1.02	.31	.98
	70–30	.07	1.11	.15	1.09	.22	1.07	.30	1.04
	90–10	.06	1.44	.13	1.43	.20	1.40	.27	1.37
	95– 5	.06	1.78	.10	1.77	.20	1.73	.25	1.71
60 or 40	50–50	.08	1.07	.17	1.05	.24	1.03	.32	.99
	70–30	.08	1.12	.15	1.11	.23	1.09	.30	1.06
	90–10	.07	1.46	.13	1.44	.20	1.42	.27	1.39
	95– 5	.06	1.80	.13	1.79	.20	1.76	.25	1.73
70 or 30	50–50	.08	1.11	.18	1.09	.27	1.05	.36	1.01
	70–30	.08	1.16	.16	1.15	.25	1.12	.33	1.08
	90–10	.07	1.51	.15	1.49	.22	1.47	.29	1.43
	95– 5	.07	1.87	.15	1.85	.22	1.81	.27	1.78
80 or 20	50–50	.11	1.20	.22	1.17	.33	1.11	.44	1.04
	70–30	.10	1.26	.20	1.23	.30	1.19	.40	1.12
	90–10	.09	1.64	.18	1.61	.25	1.58	.31	1.54
	95– 5	.09	2.03	.15	2.00	.22	1.97	.28	1.92
90 or 10	50–50	.18	1.41	.35	1.32	.53	1.15	1.00	.00
	60–40					.50	1.20	.67	.96
	70–30	.17	1.49	.31	1.41	.46	1.28	.58	1.14
	80–20					.40	1.46	.50	1.35
	90–10	.14	1.94	.25	1.89	.35	1.80	.43	1.71
	95– 5	.10	2.42	.20	2.37	.30	2.28	.40	2.16
95 or 5	50–50	.31	1.66	1.00	.00				
	60–40			.60	1.32				
	70–30	.26	1.80	.50	1.53	.76	1.02		
	80–20			.43	1.76	.60	1.49	.75	1.14
	90–10	.21	2.36	.35	2.22	.48	2.03	.58	1.82
	95– 5	.19	2.93	.31	2.80	.41	2.64	.50	2.46

TABLE FOR CONVERSION OF PERCENTAGE DIFFERENCES BETWEEN GROUPS INTO COEFFICIENTS OF TETRACHORIC CORRELATIONS.—*Continued*

Percentage of whole group above criterion	Relative sizes of compared subgroups	Difference of 25 per cent		Difference of 30 per cent		Difference of 35 per cent		Difference of 40 per cent	
		r_t	$PE(\sqrt{N})$	r_t	$PE(\sqrt{N})$	r_t	$PE(\sqrt{N})$	r_t	$PE(\sqrt{N})$
50	50–50	.39	.94	.46	.89	.53	.84	.60	.77
	70–30	.38	1.00	.45	.95	.51	.90	.58	.83
	90–10	.34	1.32	.40	1.28	.49	1.19	.55	1.12
	95– 5	.30	1.67	.40	1.58	.50	1.46	.55	1.38
60 or 40	50–50	.40	.95	.47	.90	.53	.85	.61	.77
	70–30	.39	1.01	.45	.96	.51	.91	.58	.84
	90–10	.33	1.35	.40	1.30	.45	1.25	.50	1.20
	95– 5	.30	1.69	.35	1.66	.40	1.60	.47	1.52
70 or 30	50–50	.44	.96	.50	.91	.60	.81	.67	.73
	60–40							.64	.78
	70–30	.40	1.04	.48	.97	.55	.91	.61	.84
	80–20							.57	.96
	90–10	.35	1.39	.41	1.33	.46	1.28	.51	1.23
	95– 5	.33	1.73	.40	1.66	.43	1.63	.50	1.53
80 or 20	50–50	.55	.94	.66	.81	.78	.62	1.00	.00
	60–40			.62	.87	.70	.76	.78	.63
	70–30	.48	1.06	.56	.97	.64	.88	.72	.76
	80–20			.52	1.10	.58	1.03	.64	.95
	90–10	.39	1.47	.45	1.46	.50	1.45	.56	1.26
	95– 5	.32	1.89	.40	1.80	.44	1.75	.50	1.66
90 or 10	60–40	1.00	.00						
	70–30	.71	.93	.82	.69				
	80–20	.60	1.20	.69	1.04	.76	.89	.83	.72
	90–10	.50	1.61	.57	1.49	.64	1.36	.70	1.23
	95– 5	.43	2.11	.50	1.99	.55	1.89	.60	1.78
95 or 5	80–20	1.00	1.00						
	90–10	.68	1.57	.75	1.36	.82	1.10	.87	.89
	95– 5	.57	2.28	.60	2.20	.71	1.83	.75	1.68

TABLE FOR CONVERSION OF PERCENTAGE DIFFERENCES BETWEEN GROUPS INTO COEFFICIENTS OF TETRACHORIC CORRELATIONS.—*Continued*

Percentage of whole group above criterion	Relative sizes of compared subgroups	Difference of 45 per cent		Difference of 50 per cent		Difference of 55 per cent		Difference of 60 per cent	
		r_t	PE($\sqrt{N}$)	r_t	PE($\sqrt{N}$)	r_t	PE($\sqrt{N}$)	r_t	PE($\sqrt{N}$)
50	50–50	.65	.72	.71	.64	.77	.56	.82	.48
	70–30	.64	.77	.70	.69	.77	.59	.84	.47
	80–20							.90	.47
	90–10	.61	1.04	.71	.88	.80	.70		
	95– 5	.65	1.21	.75	.99				
60 or 40	50–50	.67	.70	.74	.61	.80	.52	.84	.45
	60–40			.72	.65	.77	.57	.82	.49
	70–30	.64	.78	.70	.70	.75	.63	.80	.55
	80–20			.68	.79	.73	.71	.79	.61
	90–10	.58	1.10	.65	.99	.73	.85	.79	.73
	95– 5	.50	1.48	.70	1.12	.77	.96	.90	.55
70 or 30	50–50	.75	.62	.82	.50	.89	.36	1.00	.00
	60–40	.71	.69	.77	.60	.83	.49	.87	.41
	70–30	.67	.77	.72	.70	.78	.60	.82	.53
	80–20	.63	.89	.69	.80	.73	.74	.79	.64
	90–10	.59	1.12	.64	1.05	.70	.94	.76	.83
	95– 5	.53	1.49	.58	1.41	.65	1.28	.70	1.17
80 or 20	60–40	.86	.47	1.00	.00				
	65–35			.87	.45				
	70–30	.77	.67	.83	.55	.88	.44	.94	.27
	80–20	.70	.86	.75	.77	.80	.67	.84	.58
	90–10	.60	1.20	.67	1.08	.71	1.00	.76	.89
	95– 5	.54	1.60	.58	1.53	.65	1.48	.70	1.27
90 or 10	80–20	.91	.47	1.00	.00				
	85–15			.87	.64				
	90–10	.75	1.10	.79	.99	.83	.86	.87	.72
	95– 5	.66	1.62	.70	1.51	.74	1.39	.78	1.25
95 or 5	90–10	.92	.64	1.00	.00				
	95– 5	.79	1.50	.82	1.36	.86	1.16	.89	.98

TABLE FOR CONVERSION OF PERCENTAGE DIFFERENCES BETWEEN GROUPS INTO COEFFICIENTS OF TETRACHORIC CORRELATIONS.—*Continued*

Percentage of whole group above criterion	Relative sizes of compared subgroups	Difference of 65 per cent		Difference of 70 per cent		Difference of 75 per cent		Difference of 80 per cent	
		r_t	PE($\sqrt{N}$)	r_t	PE($\sqrt{N}$)	r_t	PR($\sqrt{N}$)	r_t	PE($\sqrt{N}$)
50	50–50	.85	.42	.90	.32	.92	.28	.94	.23
	60–40			.90	.33	.93	.26	.97	.14
	70–30	.90	.34	.94	.24				
	75–25	.95	.22						
60 or 40	50–50	.88	.37	.92	.28	.96	.17	1.00	.00
	60–40	.86	.42	.90	.33	.92	.28	.95	.20
	70–30	.84	.47	.89	.37	.94	.24	.95	.21
	80–20	.85	.49	.91	.35	1.00	.00		
	85–15			.98	.13				
	90–10	.93	.35						
70 or 30	60–40	.93	.27	.96	.18	1.00	.00		
	70–30	.86	.45	.90	.36	.93	.28	.96	.19
	80–20	.83	.55	.87	.46	.90	.39	.94	.27
	90–10	.79	.76	.86	.58	.90	.46		
	95– 5	.80	.91	.90	.57				
80 or 20	70–30	.97	.17						
	75–25			.94	.28				
	80–20	.88	.48	.92	.36	.94	.29	.96	.22
	90–10	.79	.82	.84	.69	.89	.54	.93	.39
	95– 5	.74	1.16	.78	1.05	.81	.95	.90	.62
90 or 10	90–10	.90	.60	.93	.47	.95	.37	.97	.26
	95– 5	.80	1.18	.85	.98	.89	.79	.91	.69
95 or 5	95– 5	.93	.72	.94	.64	.95	.57	.97	.39

Within the limits of this table (*i.e.*, where neither dichotomy is more extreme than 95 to 5):

A difference of 85 per cent would represent an r_t of .94 or higher
A difference of 90 per cent would represent an r_t of .98 or higher
A difference of 95 per cent would represent an r_t of .99 or higher

Percentage Difference.—This is the difference between the percentages (above the criterion) of each of the two compared subgroups, where the total of each subgroup is considered to be 100 per cent.

Percentage of Whole Group above Criterion.—This percentage must be calculated by multiplying the percentage (above the criterion) of each subgroup by the ratio of each subgroup total to the total of the whole group and then adding these two products together. (These ratios are also used in the next column.)

Relative Sizes of Compared Subgroups.—This is simply the ratio of each subgroup total to the total of the whole group.

The Tetrachoric Coefficient, r_t.—The sign of r_t is determined by inspection, for the group with the larger percentage above the criterion will obviously be positively correlated with the variable in question. The *rl*'s are omitted where they are easily interpolated or where percentage differences are impossible with the dichotomies used. With the more extreme dichotomies, the r_t's given here may be in error by from one to three hundredths, but such errors, at least for groups of the size usually treated, will always be within the limits of one probable error.

PE($\sqrt{N}$).—To find the probable error for each r_t, the tabled entry for PE($\sqrt{N}$) must be divided by the square root of the total number of individuals in the whole group considered. The ratio of r_t to the PE so obtained not only gives the reliability of the correlation but also indicates quite accurately the reliability of the percentage difference.

Interpolation.—Linear interpolation throughout the table will introduce slight errors, rarely approaching one probable error in size. For many purposes, the coefficients obtained without interpolation will be found quite accurate enough.

Examples.—Suppose that seventy per cent of the two hundred children of "bright" parents are found to have IQ's of 115 or over, while only thirty per cent of the fourteen hundred children of "average" and "dull" parents are found to have IQ's of 115 or over. The *percentage difference* between these two groups is seventy per cent minus thirty per cent, or forty per cent. The *relative sizes of the compared subgroups are* $^{200}/_{1600}$, or .125, and $^{1400}/_{1600}$, or .875. (The *percentage of the whole group above the criterion is* .125 × 70 plus .875 × 30, or 35%.) The r_t tabled most closely to these figures is that for a percentage above the criterion of forty per cent and a subgroup ratio of 90-10. This coefficient, without interpolation, then, is .50 with a PE of .03. Interpolation raises this to .52 with a PE of .03.

THE RELATION OF COLLEGE ACHIEVEMENT TESTS TO GRADES AND TO INTELLIGENCE

MELVIN G. RIGG

Oklahoma Agricultural and Mechanical College, Stillwater, Okla.

There has been for many years a widespread skepticism concerning the validity of course grades. Although this attitude has included colleges as well as elementary and secondary schools, it has been only since the development of objective standardized tests at the college level that any other measures have been available with which to compare college grades. The present paper is a report of such a comparison in seven successive graduating classes of a small college for men. Correlations have been made between achievement scores and grades, and both of these measures have been correlated with the intelligence scores previously made by the same men when they entered the institution as freshmen. These correlations are presented in the table.

TABLE SHOWING CORRELATIONS

Classes	Intelligence and first semester's scholarship		Intelligence and four years' scholarship		Intelligence and achievement scores		Achievement scores and four years' scholarship	
	N	*r*	*N*	*r*	*N*	*r*	*N*	*r*
1	112	.50	32	.41	30	.49	37	.46
2	105	.60	37	.53	35	.63	39	.52
3	85	.36	36	.32	31	.64	36	.57
4	82	.63	27	.41	27	.43	24	.36
5	77	.38	36	.46	25	.46	27	.47
6	77	.64	31	.67	25	.69	25	.47
7	77	.54	29	.22	32	.38	29	.45
Average.....	...	.52	...	.43	...	.53	...	.47
Composite...	...	...	228	.42	205	.56	217	.46

The intelligence test used was the American Council Psychological Examination. This was taken by the students as they entered the college as freshmen. A different form was used each year.

These intelligence scores were correlated, first, with grade averages for the first semester of the freshman year; and, second, with grade

averages for the entire four years of college. It will be noticed that the size of each class diminishes greatly from the first to the fourth year.

For the composite correlation in which all seven graduating classes were combined, national percentile ratings were used.

The objective achievement test, used for all classes except the third, was that devised under the direction of the Carnegie Foundation for their Pennsylvania study of 1928. After this study was completed, the remaining test booklets were sold to the college and given eventually to six graduating classes in May of their senior year. This test was identical from year to year. It was given in four sessions of three hours each. It covered practically every subject taught in the arts college, and contained something like thirty-five hundred items. The data from a test as comprehensive as this can not fail to have a considerable significance.

Class Three escaped this test, but these students had taken in their junior year the "Carnegie Achievement Tests for Sophomores." The common subjects[1] of this battery formed a convenient substitute for the 1928 test, since they consisted of essentially the same sort of material, somewhat less extensively covered.

For the composite correlations, involving all seven classes, the scores on the 1928 test and the scores made on these common subjects by Class Three were reduced to standard scores. Since these composite correlations, secured only with a considerable amount of extra labor, are very close to the averages of the seven coefficients in each column, the question arises whether the additional labor was justified.

SIGNIFICANCE OF THE FINDINGS

The correlations between intelligence and first semester's scholarship are higher than those between intelligence and four years' scholarship. This fact is a function of the narrowing range of abilities as the class proceeds toward graduation, since the poorer students tend to drop out. The differences in the N's for the two columns are significant.

Of special note are the low correlations between the grades and the objective achievement scores. Since these two are supposedly measures of the same thing, and since the achievement tests are so

[1] These common subjects were: English, usage and literature; mathematics; general culture, including general science, foreign literature, fine arts, and social studies. The tests were all objective.

comprehensive they can not easily be laughed out of court, it may be asked just what significance can be attached to teachers' marks.

This study was conducted in a liberal arts college where traditional practices prevail. The examinations are mostly of the essay type. The writer believes that grades in such a college are primarily measures of coöperation with the teacher, something which is closely akin to mastery of the subject-matter, but not identical with it. In order to secure high grades several factors other than this mastery are important:

(1) Doing just what the assignment calls for,

(2) Doing work on time,

(3) Giving the teacher back his own ideas,

(4) Presenting the work the way the teacher wants it; for instance, some teachers want voluminous examination papers; others prefer short and concise answers,

(5) Diplomacy, (the students call it politicking),

(6) Regularity of attendance, at least on strategic days when tests are scheduled or papers are due.

The student who is more independent, both as to his ideas and as to his working schedule, or the student who is less tactful, may in the end have a good mastery of the subject-matter, but his grades will be considerably lower.

It was interesting to look over the elections to Phi Beta Kappa from these seven classes. In about three-fourths of the cases these students scored well up to the top of the list on the achievement test. But in the other fourth of the cases the achievement scores were surprisingly low, even less than the average of the class. These persons had been relatively more successful in securing marks than in mastering content.

On the other hand, in every class there were one or two students with very mediocre standing in the registrar's office who ranked up toward the top of the class, often in first or second place, on the achievement test. These persons had evidently obtained an excellent education with a minimum of coöperation with the teaching staff.

Two students illustrate these extremes. M. was an expert in the technique of obtaining high grades. Everything was done on schedule time, with the instructor's views neatly paraphrased. But he was far below the average of the class on the achievement test. W., on the other hand, had never been suspected of much in the way of intellectual attainments until he came out second on the achieve-

ment test, with a score several hundred points above the class average. He had always been more or less indifferent to assignments, and had an almost straight C average. The writer had both of these men in his class during the second semester of their senior year, when the achievement test was given, and was tremendously impressed with both of the unexpected scores. The cases were discussed at some length in faculty circles. And yet, at the end of the semester, the writer gave to M. his accustomed A and to W. his accustomed C. The former had met the specific requirements of the course in a highly satisfactory way; the latter had met these but partially.

The correlations between the intelligence scores and the achievement scores are in general the highest of those computed. Presumably this relationship is less affected by extraneous factors.

Perhaps the most important aspect of this paper is to raise the question whether, since course grades do not agree very closely with achievement scores, bachelor's degrees should continue to be granted on the basis of course grades alone. It would seem to the writer quite defensible to base academic distinctions upon a combination of course grades and a system of carefully validated examinations.